OM protein - protein search, using sw model

Run on: August 22, 2004, 10:46:57 ; Search time 123 Seconds
(without alignments)
20.674 Million cell updates/sec

Title: US-10-706-475-10
Perfect score: 43
Sequence: 1 SLLMWITQX 9

Scoring table: BLOSUM62XX
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	100.0	9	4	AAB85306	Aab85306 HLA-A2 bi
2	43	100.0	10	4	AAB85314	Aab85314 HLA-A2 bi
3	42	97.7	9	2	AAW62586	Aaw62586 Cancer as
4	42	97.7	9	2	AAY01762	Aay01762 Exemplary
5	42	97.7	9	3	AAY52432	Aay52432 Human tum
6	42	97.7	9	3	AAB22791	Aab22791 NY-ESO-1
7	42	97.7	9	3	AAY78470	Aay78470 NY-ESO-1
8	42	97.7	9	3	AAB02631	Aab02631 Tumour as
9	42	97.7	9	3	AAB08703	Aab08703 Antigenic
10	42	97.7	9	4	AAE02120	Aae02120 NY-ESO-1
11	42	97.7	9	4	AAB69948	Aab69948 Human NY-
12	42	97.7	9	4	AAG67166	Aag67166 Cancer te
13	42	97.7	9	4	AAU01537	Aau01537 Cytolytic
14	42	97.7	9	4	AAB31329	Aab31329 Exemplary
15	42	97.7	9	4	AAB85299	Aab85299 HLA-A2 bi
16	42	97.7	9	4	AAB85315	Aab85315 HLA-A2 bi
17	42	97.7	9	4	AAB85304	Aab85304 HLA-A2 bi
18	42	97.7	9	4	AAB85303	Aab85303 HLA-A2 bi
19	42	97.7	9	4	AAB85305	Aab85305 HLA-A2 bi
20	42	97.7	9	4	AAB85308	Aab85308 HLA-A2 bi
21	42	97.7	9	4	AAB82017	Aab82017 HLA- bind
22	42	97.7	9	4	AAE06850	Aae06850 Human NY-
23	42	97.7	9	5	AAE26809	Aae26809 Human HLA
24	42	97.7	9	5	AAE26808	Aae26808 Human HLA
25	42	97.7	9	5	AAO21430	Aao21430 HLA-A2 re
26	42	97.7	9	6	ABP74313	Abp74313 Human NY-
27	42	97.7	9	6	ABU64813	Abu64813 Human NY-
28	42	97.7	9	6	ADA19553	Ada19553 Human can
29	42	97.7	9	7	ADC09172	Adc09172 Epitope w
30	42	97.7	9	7	ADD35560	Add35560 Human NY-
31	42	97.7	10	2	AAY06006	Aay06006 Human can
32	42	97.7	10	4	AAB85310	Aab85310 HLA-A2 bi
33	42	97.7	10	4	AAB85313	Aab85313 HLA-A2 bi
34	42	97.7	10	4	AAB85312	Aab85312 HLA-A2 bi
35	42	97.7	10	4	AAB85309	Aab85309 HLA-A2 bi
36	42	97.7	10	4	AAB85311	Aab85311 HLA-A2 bi
37	42	97.7	10	4	AAB85307	Aab85307 HLA-A2 bi
38	42	97.7	11	2	AAW62585	Aaw62585 Cancer as
39	42	97.7	11	2	AAY01761	Aay01761 Exemplary
40	42	97.7	11	3	AAY52431	Aay52431 Human tum
41	42	97.7	11	3	AAB22790	Aab22790 NY-ESO-1
42	42	97.7	11	3	AAY78469	Aay78469 NY-ESO-1
43	42	97.7	11	3	AAB02630	Aab02630 Tumour as
44	42	97.7	11	3	AAB08702	Aab08702 Antigenic
45	42	97.7	11	4	AAE02119	Aae02119 NY-ESO-1
46	42	97.7	11	4	AAB69947	Aab69947 Human NY-
47	42	97.7	11	4	AAG67165	Aag67165 Cancer te
48	42	97.7	11	4	AAU01536	Aau01536 Cytolytic
49	42	97.7	11	4	AAB31327	Aab31327 Exemplary
50	42	97.7	11	4	AAB31328	Aab31328 Exemplary

ALIGNMENTS

RESULT 1
AAB85306
ID AAB85306 standard; peptide; 9 AA.
XX
AC AAB85306;
XX
DT 17-SEP-2001 (first entry)
XX
DE HLA-A2 binding NY-ESO-1 peptide.
XX
KW NY-ESO-1; human leukocyte antigen; HLA; lysis; cytolytic T cell; CTL;
KW HLA-A2; T-cell sorter; tumor; immune tetramer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 9
FT /label= Ala, Val, Leu, Ile, Pro, Phe, Met, Trp or Gly
FT /note= "can be any amino acid, preferably one with a non-
FT polar side chain such as those residues indicated above"
XX
PN WO200136453-A2.
XX
PD 25-MAY-2001
XX
PF 08-NOV-2000; 2000WO-US042010.
XX
PR 15-NOV-1999; 99US-00440621.
PR 25-FEB-2000; 2000US-00514036.
PR 29-SEP-2000; 2000US-00676005.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
PA (UYOX-) UNIV OXFORD.
XX
PI Valmori D, Cerottini J, Romero P, Cerundolo V;
XX
DR WPI; 2001-451454/48.
XX
PT Novel isolated NY-ESO-1 nonapeptide useful for determining if a cell
PT presents human leukocyte antigen-A2 molecule on its surface, binds to
PT human leukocyte antigen molecules and provokes lysis by cytolytic T
PT cells.
XX

```
PS   Claim 1; Page 25; 38pp; English.
XX
CC   The invention provides NY-ESO-1 peptide derivatives which bind to human
CC   leukocyte antigen (HLA) molecules and provokes lysis by cytolytic T cells
CC   (CTLs). The NY-ESO-1 nonapeptide is of formula SLLMWITQX, where X is an
CC   amino acid having an uncharged polar side chain. The NY-ESO-1 peptide
CC   derivatives are useful for determining if a cell presents an HLA-A2
CC   molecule on its surface, by contacting a sample containing the cell with
CC   the peptide or its derivative, and determining binding between them,
CC   where the binding is indicative of HLA-A2 on the surface of the cell. The
CC   NY-ESO-1 peptides and analogues are useful therapeutically, for
CC   administration to a patient who is HLA-A2 positive and expresses NY-ESO-1
CC   in connection with the pathology, as well as diagnostically, i.e. to
CC   determine if HLA-A2 positive cells are present, or if relevant CTLs are
CC   present. They are also useful for determining the presence of CTLs in a
CC   sample. The peptides are useful as T-cell sorters, when incorporated into
CC   immune tetramers. The present sequence represents a NY-ESO-1 peptide that
CC   can bind to HLA-A2 molecule
XX
SQ   Sequence 9 AA;

  Query Match             100.0%;  Score 43;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQX 9
              |||||||||
Db          1 SLLMWITQX 9


RESULT 2
AAB85314
ID   AAB85314 standard; peptide; 10 AA.
XX
AC   AAB85314;
XX
DT   17-SEP-2001  (first entry)
XX
DE   HLA-A2 binding NY-ESO-1 peptide.
XX
KW   NY-ESO-1; human leukocyte antigen; HLA; lysis; cytolytic T cell; CTL;
KW   HLA-A2; T-cell sorter; tumor; immune tetramer.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 9
FT                   /label= Ala or Cys
FT   Misc-difference 10
FT                   /label= Val, Phe, Ile or Leu
FT                   /note= "can be any amino acid, preferably one indicated
FT                   above"
XX
PN   WO200136453-A2.
XX
PD   25-MAY-2001.
XX
PF   08-NOV-2000; 2000WO-US042010.
XX
PR   15-NOV-1999;   99US-00440621.
PR   25-FEB-2000; 2000US-00514036.
PR   29-SEP-2000; 2000US-00676005.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
PA   (UYOX-) UNIV OXFORD.
XX
PI   Valmori D,  Cerottini J,  Romero P,  Cerundolo V;
XX
DR   WPI; 2001-451454/48.
XX
PT   Novel isolated NY-ESO-1 nonapeptide useful for determining if a cell
PT   presents human leukocyte antigen-A2 molecule on its surface, binds to
PT   human leukocyte antigen molecules and provokes lysis by cytolytic T
PT   cells.
XX
PS   Disclosure; Page 36; 38pp; English.
XX
CC   The invention provides NY-ESO-1 peptide derivatives which bind to human
CC   leukocyte antigen (HLA) molecules and provokes lysis by cytolytic T cells
CC   (CTLs). The NY-ESO-1 nonapeptide is of formula SLLMWITQX, where X is an
CC   amino acid having an uncharged polar side chain. The NY-ESO-1 peptide
CC   derivatives are useful for determining if a cell presents an HLA-A2
CC   molecule on its surface, by contacting a sample containing the cell with
CC   the peptide or its derivative, and determining binding between them,
CC   where the binding is indicative of HLA-A2 on the surface of the cell. The
CC   NY-ESO-1 peptides and analogues are useful therapeutically, for
CC   administration to a patient who is HLA-A2 positive and expresses NY-ESO-1
CC   in connection with the pathology, as well as diagnostically, i.e. to
CC   determine if HLA-A2 positive cells are present, or if relevant CTLs are
CC   present. They are also useful for determining the presence of CTLs in a
CC   sample. The peptides are useful as T-cell sorters, when incorporated into
CC   immune tetramers. The present sequence represents a NY-ESO-1 peptide that
CC   can bind to HLA-A2 molecule
XX
SQ   Sequence 10 AA;

  Query Match             100.0%;  Score 43;  DB 4;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 0.4;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQX 9
              |||||||||
Db          1 SLLMWITQX 9


RESULT 3
AAW62586
ID   AAW62586 standard; peptide; 9 AA.
XX
AC   AAW62586;
XX
DT   17-SEP-1998  (first entry)
XX
DE   Cancer associated antigen peptide.
XX
KW   Cancer associated antigen; NY-ESO-1; regression; progression; onset;
KW   cancer; treatment; diagnosis.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9814464-A1.
XX
PD   09-APR-1998.
XX
PF   15-SEP-1997;   97WO-US016335.
XX
PR   03-OCT-1996;   96US-00725182.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Chen Y,  Scanlan M,  Gure A,  Old LJ,  Jager E,  Knuth A;
PI   Drijfhout JW;
XX
DR   WPI; 1998-286417/25.
XX
PT   New isolated cancer associated antigen - is used to develop products for
PT   the diagnosis and treatment of cancers and for monitoring cancer therapy.
XX
PS   Claim 33; Page 17; 49pp; English.
XX
CC   Peptides AAW62585-87 are derived from cancer associated antigen NY-ESO-1,
CC   and are stimulators of cytotoxic T-cells. The specification describes a
CC   method for determining regression, progression of onset of a cancerous
CC   condition, comprising monitoring a sample from a patient with the
CC   cancerous condition for a parameter selected from NY-ESO-1 protein, a
```

```
CC   peptide derived from NY-ESO-1 protein and cytolytic T cells specific for
CC   the peptide and an MHC molecule with which it non-covalently complexes.
CC   Methods for the treatment of a cancerous condition are also described.
CC   The NY-ESO-1 protein and peptides derived from it can be used for
CC   diagnosis and treatment of cancers and to monitor the efficacy of a
CC   therapeutic regime
XX
SQ   Sequence 9 AA;

  Query Match             97.7%;  Score 42;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              ||||||||
Db          1 SLLMWITQ 8


RESULT 4
AAY01762
ID   AAY01762 standard; peptide; 9 AA.
XX
AC   AAY01762;
XX
DT   25-JUN-1999  (first entry)
XX
DE   Exemplary antigenic peptide derived from NY-ESO-1.
XX
KW   MAGE-3; tumour associated gene; human leucocyte antigen Class II;
KW   autologous CD4+ cell; MAGE-3 related disease; cancer; melanoma;
KW   osteosarcoma; leukemia; carcinoma.
XX
OS   Homo sapiens.
XX
PN   WO9914326-A1.
XX
PD   25-MAR-1999.
XX
PF   04-SEP-1998;   98WO-US018601.
XX
PR   12-SEP-1997;   97US-00928615.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
PA   (UYVR-) UNIV VRIJE BRUSSEL.
XX
PI   Thielemans K,  Heirman C,  Corthals J,  Chaux P,  Stroobant V;
PI   Boon-Falleur T,  Van Der Bruggen P,  Luiten R;
XX
DR   WPI; 1999-244031/20.
XX
PT   Isolated peptides that bind to human leucocyte antigen class II
PT   molecules.
XX
PS   Disclosure; Page 29; 88pp; English.
XX
CC   The present sequence represents an exemplary tumour associated peptide
CC   antigen. The specification describes a MAGE-3 tumour associated gene.
CC   Peptides (AAY01721-25) that bind human leucocyte antigen (HLA) Class II
CC   molecules can be derived from the MAGE-3 protein. These peptides and
CC   autologous CD4+ cells that bind to a complex of MAGE-3 peptide and HLA
CC   Class II, are used to treat MAGE-3 related diseases, particularly cancers
CC   (e.g. melanoma, osteosarcoma, leukemia and various forms of carcinoma).
CC   The peptides are also used to produce specific antibodies. Detection of
CC   of the peptides, e.g. in binding assays, particularly with antibodies, is
CC   used for diagnosis of such diseases
XX
SQ   Sequence 9 AA;

  Query Match             97.7%;  Score 42;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              ||||||||
Db          1 SLLMWITQ 8


RESULT 5
AAY52432
ID   AAY52432 standard; peptide; 9 AA.
XX
AC   AAY52432;
XX
DT   15-FEB-2000  (first entry)
XX
DE   Human tumour antigen NY-ESO-1 peptide #5.
XX
KW   Cancer; tumour; antigen; MHC; major histocompatility complex; Class I;
KW   T-cell; cytotoxic; stimulation; proliferation; treatment; diagnosis;
KW   prevention; melanoma; breast cancer; ovarian cancer; prostate cancer;
KW   hepatoma; thyroid cancer; bladder cancer; lung cancer; lymphoma.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9953938-A1.
XX
PD   28-OCT-1999.
XX
PF   24-MAR-1999;   99WO-US006875.
XX
PR   17-APR-1998;   98US-00062422.
PR   02-OCT-1998;   98US-00165546.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Stockert E,  Jager E,  Chen Y,  Scanlan M,  Alexander K,  Old LJ;
PI   Gure A,  Ritter G;
XX
DR   WPI; 2000-038483/03.
XX
PT   Novel peptides which bind to MHC class I and MHC class II molecules,
PT   useful for therapeutic and diagnostic purposes.
XX
PS   Claim 60; Page 18; 49pp; English.
XX
CC   Peptides #4-#7 (AAY52431-Y52434) are peptides derived from the human
CC   tumour antigen, NY-ESO-1 (AAY52430) which contain the motif LLMWIT
CC   (AAY52441). These sequences can bind to MHC (major histocompatibility
CC   Class I HLA-A2 molecules, thereby stimulating proliferation of cytotoxic
CC   T-cells. cDNA encoding NY-ESO-1 was initially isolated from an oesophagus
CC   squamous cell cancer cDNA library. Tissue localisation studies revealed
CC   it to be expressed at high levels in normal ovary and testis but not in
CC   normal colon, kidney, liver, brain, oesophagus and skin. It was expressed
CC   in certain tumours and tumour cell lines with some degree of frequency -
CC   these included melanoma specimens and cell lines, and breast and bladder
CC   cancer specimens, with expression in other tumour types being sporadic.
CC   These NY-ESO-1-derived peptides may be used in methods and compositions
CC   used for the treatment, diagnosis and prevention of cancers (such as
CC   melanoma, breast cancer, prostate cancer, lung cancer, hepatoma, ovarian
CC   cancer, thyroid cancer, bladder cancer, or lymphoma) and to stimulate the
CC   proliferation of T cells
XX
SQ   Sequence 9 AA;

  Query Match             97.7%;  Score 42;  DB 3;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              ||||||||
Db          1 SLLMWITQ 8


RESULT 6
AAB22791
```

```
ID   AAB22791 standard; peptide; 9 AA.
XX
AC   AAB22791;
XX
DT   22-DEC-2000  (first entry)
XX
DE   NY-ESO-1 peptide epitope, SEQ ID NO:2.
XX
KW   NY-ESO-1; epitope; CTL response; cytotoxic T lymphocyte; vaccine;
KW   immunogenic; adjuvant coadministration; microbial infection;
KW   tuberculosis; HIV; hepatitis B virus; hepatitis C virus; cancer.
XX
OS   Unidentified.
XX
PN   WO200048630-A1.
XX
PD   24-AUG-2000.
XX
PF   17-FEB-2000; 2000WO-AU000110.
XX
PR   17-FEB-1999;   99AU-00008735.
PR   27-JUL-1999;   99AU-00001861.
XX
PA   (CSLC-) CSL LTD.
XX
PI   Cox JC,  Drane DP;
XX
DR   WPI; 2000-571930/53.
XX
PT   Immunogenic complexes comprising negatively charged organic carrier
PT   adjuvants and positively charged antigens for use as vaccines against
PT   microbial infection and cancer.
XX
PS   Example 4; Fig 5c; 111pp; English.
XX
CC   The invention relates to a novel immunogenic complex comprising a charged
CC   organic carrier and a charged antigen which are electrostatically
CC   associated. The complex induces a cytotoxic T lymphocyte (CTL) response.
CC   The complex and/or vaccine can be used to treat a disease in a mammal,
CC   where the complex/vaccine elicits, induces or otherwise facilitates an
CC   immune response which inhibits, halts, delays or prevents the onset or
CC   progression of the disease condition. In particular, the disease is a
CC   condition resulting form a microbial infection or cancer. Microbial
CC   infections which may be treated using the immunogenic complex include
CC   human immunodeficiency virus (HIV), hepatitis B, hepatitis C,
CC   tuberculosis or a parasitic condition, and cancers which may be treated
CC   include melanoma, prostate cancer or breast cancer. The complexes and
CC   vaccines simultaneously co-deliver antigen and adjuvant to the same
CC   antigen presenting cell, which is often essential for induction of
CC   appropriate immune responses. Sequences AAB22790-B22791 represent peptide
CC   epitopes of the positively charged protein NY-ESO-1 used in an
CC   exemplification of the invention
XX
SQ   Sequence 9 AA;

  Query Match             97.7%;  Score 42;  DB 3;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              ||||||||
Db          1 SLLMWITQ 8


RESULT 7
AAY78470
ID   AAY78470 standard; peptide; 9 AA.
XX
AC   AAY78470;
XX
DT   10-MAY-2000  (first entry)
XX
DE   NY-ESO-1 derived peptide #2.
XX
KW   Cancer; SSX family; SSX-1; SSX-2; SSX-3; SSX-4; SSX-5; NY-ESO-1;
KW   HLA binding; human leukocyte antigen; cytolytic T cell; CTL; cytostatic;
KW   melanoma; synovial sarcoma.
XX
OS   Homo sapiens.
XX
PN   WO200000824-A1.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US014493.
XX
PR   26-JUN-1998;   98US-00105839.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Tureci O,  Sahin U,  Pfreundschuh M,  Rammensee G,  Stevanovic S;
PI   Chen Y,  Gure A,  Old LJ;
XX
DR   WPI; 2000-170933/15.
XX
PT   Determining the possible presence of breast, endometrial, colorectal,
PT   lung, bladder or head-neck cancer.
XX
PS   Example 12; Page 21; 40pp; English.
XX
CC   A method has been developed for determining the possible presence of a
CC   cancer, which is not melanoma or synovial sarcoma. The method comprises
CC   assaying a sample taken from the subject to determine the expression of
CC   an SSX gene, and determining the expression as a determination of the
CC   possible presence of cancer. Expression of SSX1 gene indicates possible
CC   presence of breast, endometrial, colorectal, lung, bladder or head-neck
CC   cancer. These cancers are also detected by SSX2 and SSX4 gene expression.
CC   SSX2 gene expression additionally indicates possible presence of
CC   lymphoma, renal cell cancer, glioma and prostate cancer. Expression of
CC   SSX4 gene also indicates possible presence of ovarian or stomach cancer.
CC   SSX5 gene expression indicates the same cancers as SSX1, except breast
CC   cancer. Determining expression of SSX gene can be used to monitor
CC   progress of melanoma or synovial sarcoma, which is not cancer. The SSX-
CC   derived peptide complex stimulates proliferation of cytolytic T cells.
CC   This is useful for treating cancer, especially melanoma. AAY78464 to
CC   AAY78468 represent specifically claimed HLA binding peptides for use in
CC   the method of the invention. AAZ88452 to AAZ88465 represent PCR primers
CC   used in the isolation of SSX genes in the exemplification of the present
CC   invention. AAY78469 to AAY78500, and AAY79684 to AAY79762 represent
CC   peptides derived from SSX proteins or NY-ESO-1, which are used in the
CC   exemplification of the present invention
XX
SQ   Sequence 9 AA;

  Query Match             97.7%;  Score 42;  DB 3;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              ||||||||
Db          1 SLLMWITQ 8


RESULT 8
AAB02631
ID   AAB02631 standard; peptide; 9 AA.
XX
AC   AAB02631;
XX
DT   18-AUG-2000  (first entry)
XX
DE   Tumour associated peptide antigen from NY-ESO-1 #2.
XX
KW   MAGE-A3; HLA class II; human leukocyte antigen; antibody; vaccine;
KW   cancer; human; tumour; tumour associated gene product.
XX
```

OS Homo sapiens.
XX
PN WO200020581-A1.
XX
PD 13-APR-2000.
XX
PF 15-SEP-1999; 99WO-US021230.
XX
PR 05-OCT-1998; 98US-00166448.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
PA (UYVR-) UNIV VRIJE BRUSSEL.
XX
PI Chaux P, Stroobant V, Boon-Falleur T, Van Der Bruggen P;
PI Schultz ES, Van Snick J, Lethe B, Thielemans K, Corthals J;
PI Heirman C;
XX
DR WPI; 2000-317713/27.
XX
PT New MAGE-A3 class II binding peptides, useful to diagnose and treat
PT tumors, are fragments of MAGE-A3 which bind to and are presented to T
PT lymphocytes by human leukocyte antigen class II molecules.
XX
PS Disclosure; Page 33; 119pp; English.
XX
CC The present invention relates to MAGE-A3 (tumour associated gene product)
CC human leukocyte antigen (HLA) class II-binding peptides (see AAB02566-
CC B02595, and AAB02633-B02637). These peptides are presented to T cells in
CC the context of HLA class II molecules. The peptides stimulate the
CC activity and proliferation of CD4+ T lymphocytes. The invention also
CC includes nucleotide sequences encoding MAGE-3A peptides (see AAA37928 and
CC AAA37938-A37940). The peptides and nucleotide sequences can be used to
CC create antibodies against the MAGE-A3 peptides, the antibodies, peptides
CC and nucleotide sequences can be used to create a vaccine. The peptides
CC are used to diagnose or treat a disorder characterized by expression of
CC MAGE-3, particularly cancer. The methods can also be used in the
CC diagnosis of disorders associated with MAGE-3 expression. Included in the
CC invention are other human tumour antigens (see AAB02596-B02637), and PCR
CC primers used in the course of the invention (see AAA37929-A37937 and
CC AAA37941-A37942)
XX
SQ Sequence 9 AA;

Query Match 97.7%; Score 42; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLMWITQ 8
||||||||
Db 1 SLLMWITQ 8

RESULT 9
AAB08703
ID AAB08703 standard; peptide; 9 AA.
XX
AC AAB08703;
XX
DT 02-JAN-2001 (first entry)
XX
DE Antigenic peptide from tumour rejection antigen NY-ESO-1.
XX
KW EphA3; HLA class II-binding peptide; human leukocyte antigen; antigen;
KW CD4+ T lymphocyte; tumour associated gene; vaccine.
XX
OS Homo sapiens.
XX
PN WO200050589-A1.
XX
PD 31-AUG-2000.
XX
PF 18-FEB-2000; 2000WO-US004326.
XX
PR 22-FEB-1999; 99US-0121170P.
PR 08-OCT-1999; 99US-0158566P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chiari R, Coulie P, Boon-Falleur T;
XX
DR WPI; 2000-572089/53.
XX
PT Novel tyrosine kinase receptor, EphA3 human leukocyte antigen (HLA) class
PT II binding peptide and nucleic acid encoding the receptor, useful for
PT diagnosing and treating conditions characterized by expression of EphA3
PT gene.
XX
PS Disclosure; Page 36; 107pp; English.
XX
CC AAB08668-B08704 represent antigenic peptides characteristic of tumours.
CC The peptides may be combined in vaccines with a human EphA3 HLA (human
CC leukocyte antigen) class II-binding peptide. EphA3 antigens, when
CC presented by an antigen presenting cell having a HLA class II molecule,
CC effectively induce activation and proliferation of CD4+ T lymphocytes.
CC EphA3 is a tumour associated gene. EphA3 HLA binding peptides are used
CC for selectively enriching a population of T lymphocytes. The peptides are
CC also used for diagnosing a disorder characterized by EphA3 or EphA3 HLA
CC binding peptide expression. The peptides are also used to treat a
CC disorder characterized by EphA3 expression. The EphA3 binding peptides
CC are useful in producing vaccines and antibody
XX
SQ Sequence 9 AA;

Query Match 97.7%; Score 42; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLMWITQ 8
||||||||
Db 1 SLLMWITQ 8

RESULT 10
AAE02120
ID AAE02120 standard; peptide; 9 AA.
XX
AC AAE02120;
XX
DT 31-JUL-2001 (first entry)
XX
DE NY-ESO-1 human leukocyte antigen-A2-binding peptide #2.
XX
KW Human; cytostatic; immunogen; NY-ESO-1; human leukocyte antigen; HLA;
KW CD8; cytotoxic T lymphocyte; cancer; carcinoma; melanoma; myeloma;
KW brain tumour; sarcoma; vaccine; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200129220-A2.
XX
PD 26-APR-2001.
XX
PF 19-OCT-2000; 2000WO-US028852.
XX
PR 19-OCT-1999; 99US-0160374P.
PR 01-FEB-2000; 2000US-0179570P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Heidecker L, Van Den Eynde B, Boon-Falleur T, Brasseur F;
XX
DR WPI; 2001-328498/34.
XX
PT New antigenic peptides derived from MAGE-A12 polypeptides, useful for
PT diagnosis and treatment of cancer, such as bladder, lung, breast, brain,
PT prostate and renal carcinomas.

```
XX
PS   Disclosure; Page 21; 69pp; English.
XX
CC   The patent discloses antigenic peptides derived from MAGE-A12 protein and
CC   presented by human leukocyte antigens (HLAs). These antigenic peptides
CC   when presented by an antigen presenting cell having a HLA class I
CC   molecule, effectively induce the activation and proliferation of CD8+
CC   cytotoxic T lymphocytes (CTLs). MAGE-A12 is useful for treating a subject
CC   having a disorder characterised by expression of MAGE-A12. The protein
CC   microarray comprising MAGE-A12 is useful for diagnosing a disorder,
CC   especially cancer, by determining the binding of an antibody, T
CC   lymphocytes or a HLA molecule isolated from the subject suspected of
CC   having the disorder characterised by the expression of MAGE-A12. MAGE-A12
CC   is useful for treating cancers, including bladder carcinomas, melanomas,
CC   oesophageal, lung, head and neck, breast, colorectal carcinomas,
CC   myelomas, brain tumours, sarcomas, prostate and renal carcinomas and to
CC   produce antibodies. MAGE-A12 antibodies are useful for diagnosing
CC   disorders characterised by expression of MAGE-A12 immunogenic
CC   polypeptide. These MAGE-A12 peptides are used as vaccines. They are also
CC   used in gene therapy. The present sequence is an antigenic peptide
CC   derived from NY-ESO-1. This peptide which is characteristic of tumours is
CC   presented by HLA-A2 MHC (major histocompatibility complex) and is
CC   recognised by CTLs
XX
SQ   Sequence 9 AA;

  Query Match             97.7%;  Score 42;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              ||||||||
Db          1 SLLMWITQ 8


RESULT 11
AAB69948
ID   AAB69948 standard; peptide; 9 AA.
XX
AC   AAB69948;
XX
DT   27-APR-2001  (first entry)
XX
DE   Human NY-ESO-1 CTL stimulating peptide #2.
XX
KW   Human; NY-ESO-1; HLA; human leukocyte antigen; CTL; cytotoxic T cell;
KW   HLA-A2; HLA-DR53; melanoma; adenocarcinoma; bladder carcinoma;
KW   non-small cell lung carcinoma; tumour status determination.
XX
OS   Homo sapiens.
XX
PN   WO200107917-A1.
XX
PD   01-FEB-2001.
XX
PF   14-JUL-2000; 2000WO-US019220.
XX
PR   23-JUL-1999;   99US-00359503.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
PA   (SLOK ) SLOAN KETTERING INST CANCER RES.
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Jager E,  Stockert E,  Old LJ,  Knuth A,  Chen Y,  Scanlan M;
XX
DR   WPI; 2001-182822/18.
XX
PT   Method useful for determining the status (e.g. progression, regression or
PT   stability of the disease) of a cancerous condition, involves determining
PT   the levels of NY-ESO-1 specific antibodies in a sample taken from a
PT   patient.
XX
PS   Example 13; Page 24; 50pp; English.
XX
CC   The present sequence is given in a specification relating to a method for
CC   determining the status of a cancerous condition in a patient with a
CC   tumour that expresses NY-ESO-1. The method comprises assaying a sample
CC   taken from the patient for antibodies that specifically bind to the NY-
CC   ESO-1 and comparing the value obtained to a prior value obtained from
CC   assay of a prior sample taken from the patient. Any difference between
CC   the values is indicative of a change in status of the cancerous
CC   condition. The method is useful for determining whether a cancerous
CC   condition is progressing, regressing or remaining stable, in particular
CC   in patients receiving treatment for a melanoma, adenocarcinoma, non-small
CC   cell lung carcinoma or bladder carcinoma
XX
SQ   Sequence 9 AA;

  Query Match             97.7%;  Score 42;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              ||||||||
Db          1 SLLMWITQ 8


RESULT 12
AAG67166
ID   AAG67166 standard; peptide; 9 AA.
XX
AC   AAG67166;
XX
DT   13-NOV-2001  (first entry)
XX
DE   Cancer testis tumour antigen NY-ESO-1 derived CTL-stimulating peptide.
XX
KW   Cancer testis tumour antigen; NY-ESO-1; LAGE-2; human leukocyte antigen;
KW   HLA; HLA binding peptide; major histocompatibility complex; MHC; tumour;
KW   cancer; testis tumour.
XX
OS   Homo sapiens.
XX
PN   WO200162917-A1.
XX
PD   30-AUG-2001.
XX
PF   22-JAN-2001; 2001WO-US002126.
XX
PR   22-FEB-2000; 2000US-00510635.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Lethe B,  Boon-Falleur T;
XX
DR   WPI; 2001-550091/61.
XX
PT   Genomic sequences of tumor associated antigen EY-ESO-1 (LAGE-2) useful
PT   for diagnosing testicular tumors.
XX
PS   Example 12; Page 24; 50pp; English.
XX
CC   The present sequence represents a peptide which is derived from cancer
CC   testis tumour antigen NY-ESO-1 (also called LAGE-2). The peptide
CC   stimulates cytolytic T cell lines (CTLs). NY-ESO-1 is a molecule that is
CC   processed to at least one human leukocyte antigen (HLA) binding peptide,
CC   which binds to Class I and Class II major histocompatibility complex
CC   (MHC). NY-ESO-1 is expressed in tumour mRNA and in testis, but not normal
CC   colon, kidney, liver or brain tissue. The presence or level of expression
CC   of NY-ESO-1 may be assayed for the diagnosis of cancer, especially testis
CC   tumours
XX
SQ   Sequence 9 AA;

  Query Match             97.7%;  Score 42;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
```

```
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              ||||||||
Db          1 SLLMWITQ 8


RESULT 13
AAU01537
ID   AAU01537 standard; peptide; 9 AA.
XX
AC   AAU01537;
XX
DT   18-JUL-2001  (first entry)
XX
DE   Cytolytic T cell line stimulator peptide #2.
XX
KW   NY-ESO-1; human; tumour rejection antigen precursor; SSX-2; MHC Class II;
KW   major histocompatibility complex; helper T cell; HLA-DR; cancer;
KW   human leukocyte antigen-determining region; disease progression;
KW   disease regression; disease onset; body tissue; body fluid; enzyme label;
KW   radioactive label; monoclonal antibody; cytolytic T cell line.
XX
OS   Homo sapiens.
XX
PN   WO200123560-A2.
XX
PD   05-APR-2001.
XX
PF   26-SEP-2000; 2000WO-US026411.
XX
PR   29-SEP-1999;   99US-00408036.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Tureci O,  Sahin U,  Pfreundschuh M;
XX
DR   WPI; 2001-266156/27.
XX
PT   Polypeptides binding to major histocompatibility complex class II human
PT   leukocyte antigen-determining region molecule having amino acid sequence
PT   found in tumor rejection antigen precursor used for stimulating
PT   proliferation of helper T cells.
XX
PS   Example 12; Page 17; 62pp; English.
XX
CC   The sequence represents a human NY-ESO-1 tumour rejection antigen
CC   precursor fragment which efficiently stimulates cytolytic T cell lines.
CC   NY-ESO-1 and SSX-2 polypeptides, or fragments of, bind to major
CC   histocompatibility complex (MHC) Class II molecules such as human
CC   leukocyte antigen-determining region (HLA-DR) molecules and stimulate
CC   proliferation of helper T cells. The peptides can be administered to an
CC   HLA-DR positive subject in order to stimulate the helper T cells. An MHC
CC   Class II HLA-DR-NY-ESO-1/SSX-2 complex expressed on the surface of a cell
CC   or present in free form is useful for this stimulation. The nucleic acid
CC   is useful for screening for a cancerous condition, which involves
CC   contacting a subject sample to a cell line transfected with the
CC   immunoreactive cell (helper T cell), where interaction is indicative of
CC   cancer. In addition, a sample from a patient (for example, a body fluid
CC   or tissue) can be monitored for the amount of the complex present in the
CC   bloodstream. This is useful for determining regression, progression or
CC   onset of a cancerous condition. The method involves contacting the sample
CC   with a radioactive labelled or enzyme labelled monoclonal antibody which
CC   specifically binds with the complex
XX
SQ   Sequence 9 AA;

  Query Match             97.7%;  Score 42;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              ||||||||
Db          1 SLLMWITQ 8


RESULT 14
AAB31329
ID   AAB31329 standard; peptide; 9 AA.
XX
AC   AAB31329;
XX
DT   20-APR-2001  (first entry)
XX
DE   Exemplary antigen characteristic of tumours and derived from NY-ESO-1.
XX
KW   MAGE-A1; HLA; human leukocyte antigen; CD4+ T lymphocyte; cancer;
KW   MAGE-A1 HLA class II-binding protein; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200078806-A1.
XX
PD   28-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-US016287.
XX
PR   18-JUN-1999;   99US-00336091.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Van Snick J,  Lethe B,  Chaux P,  Boon-Falleur T,  Van Der Bruggen P;
XX
DR   WPI; 2001-102698/11.
XX
PT   Novel MAGE-A1 human leukocyte antigen class II peptides which bind to and
PT   are presented to the class II molecules, useful for inducing immune
PT   response and treating cancers characterized by expression of MAGE-A1.
XX
PS   Disclosure; Page 32; 78pp; English.
XX
CC   AAB31302-59 represent exemplary antigens which are characteristic of
CC   tumours. They can be used to enhance the immune response of vaccines
CC   comprising peptides derived from human MAGE-A1 HLA (human leukocyte
CC   antigen) class II-binding protein. Peptides derived from the MAGE-A1 HLA
CC   binding protein stimulate the activity and proliferation of CD4+ T
CC   lymphocytes. The MAGE-A1 HLA binding protein is useful as a diagnostic
CC   agent for diagnosing a disorder characterized by expression of MAGE-A1.
CC   The protein is used for treating a disorder characterized by expression
CC   of MAGE-A1 such as cancers e.g. melanomal, squamous cell carcinomas,
CC   colorectal carcinomas, osteosarcomas, and lymphocytic leukemias. Peptides
CC   derived from the MAGE-A1 HLA binding protein are useful in the production
CC   of anti-tumour vaccines
XX
SQ   Sequence 9 AA;

  Query Match             97.7%;  Score 42;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              ||||||||
Db          1 SLLMWITQ 8


RESULT 15
AAB85299
ID   AAB85299 standard; peptide; 9 AA.
XX
AC   AAB85299;
XX
DT   17-SEP-2001  (first entry)
XX
DE   HLA-A2 binding NY-ESO-1 peptide #2.
XX
KW   NY-ESO-1; human leukocyte antigen; HLA; lysis; cytolytic T cell; CTL;
```

KW HLA-A2; T-cell sorter; tumor; immune tetramer.
XX
OS Homo sapiens.
XX
PN WO200136453-A2.
XX
PD 25-MAY-2001.
XX
PF 08-NOV-2000; 2000WO-US042010.
XX
PR 15-NOV-1999; 99US-00440621.
PR 25-FEB-2000; 2000US-00514036.
PR 29-SEP-2000; 2000US-00676005.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
PA (UYOX-) UNIV OXFORD.
XX
PI Valmori D, Cerottini J, Romero P, Cerundolo V;
XX
DR WPI; 2001-451454/48.
XX
PT Novel isolated NY-ESO-1 nonapeptide useful for determining if a cell
PT presents human leukocyte antigen-A2 molecule on its surface, binds to
PT human leukocyte antigen molecules and provokes lysis by cytolytic T
PT cells.
XX
PS Example 1; Page 4; 38pp; English.
XX
CC The invention provides NY-ESO-1 peptide derivatives which bind to human
CC leukocyte antigen (HLA) molecules and provokes lysis by cytolytic T cells
CC (CTLs). The NY-ESO-1 nonapeptide is of formula SLLMWITQX, where X is an
CC amino acid having an uncharged polar side chain. The NY-ESO-1 peptide
CC derivatives are useful for determining if a cell presents an HLA-A2
CC molecule on its surface, by contacting a sample containing the cell with
CC the peptide or its derivative, and determining binding between them,
CC where the binding is indicative of HLA-A2 on the surface of the cell. The
CC NY-ESO-1 peptides and analogues are useful therapeutically, for
CC administration to a patient who is HLA-A2 positive and expresses NY-ESO-1
CC in connection with the pathology, as well as diagnostically, i.e. to
CC determine if HLA-A2 positive cells are present, or if relevant CTLs are
CC present. They are also useful for determining the presence of CTLs in a
CC sample. The peptides are useful as T-cell sorters, when incorporated into
CC immune tetramers. The present sequence represents a NY-ESO-1 peptide that
CC can bind to HLA-A2 molecule
XX
SQ Sequence 9 AA;

```
  Query Match             97.7%;  Score 42;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              ||||||||
Db          1 SLLMWITQ 8
```

RESULT 16
AAB85315
ID AAB85315 standard; peptide; 9 AA.
XX
AC AAB85315;
XX
DT 17-SEP-2001 (first entry)
XX
DE HLA-A2 binding NY-ESO-1 peptide analogue.
XX
KW NY-ESO-1; human leukocyte antigen; HLA; lysis; cytolytic T cell; CTL;
KW HLA-A2; T-cell sorter; tumor; immune tetramer.
XX
OS Homo sapiens.
XX
PN WO200136453-A2.
XX
PD 25-MAY-2001.
XX
PF 08-NOV-2000; 2000WO-US042010.
XX
PR 15-NOV-1999; 99US-00440621.
PR 25-FEB-2000; 2000US-00514036.
PR 29-SEP-2000; 2000US-00676005.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
PA (UYOX-) UNIV OXFORD.
XX
PI Valmori D, Cerottini J, Romero P, Cerundolo V;
XX
DR WPI; 2001-451454/48.
XX
PT Novel isolated NY-ESO-1 nonapeptide useful for determining if a cell
PT presents human leukocyte antigen-A2 molecule on its surface, binds to
PT human leukocyte antigen molecules and provokes lysis by cytolytic T
PT cells.
XX
PS Claim 6; Page 16; 38pp; English.
XX
CC The invention provides NY-ESO-1 peptide derivatives which bind to human
CC leukocyte antigen (HLA) molecules and provokes lysis by cytolytic T cells
CC (CTLs). The NY-ESO-1 nonapeptide is of formula SLLMWITQX, where X is an
CC amino acid having an uncharged polar side chain. The NY-ESO-1 peptide
CC derivatives are useful for determining if a cell presents an HLA-A2
CC molecule on its surface, by contacting a sample containing the cell with
CC the peptide or its derivative, and determining binding between them,
CC where the binding is indicative of HLA-A2 on the surface of the cell. The
CC NY-ESO-1 peptides and analogues are useful therapeutically, for
CC administration to a patient who is HLA-A2 positive and expresses NY-ESO-1
CC in connection with the pathology, as well as diagnostically, i.e. to
CC determine if HLA-A2 positive cells are present, or if relevant CTLs are
CC present. They are also useful for determining the presence of CTLs in a
CC sample. The peptides are useful as T-cell sorters, when incorporated into
CC immune tetramers. The present sequence represents a NY-ESO-1 peptide
CC analogue that can bind to HLA-A2 molecule
XX
SQ Sequence 9 AA;

```
  Query Match             97.7%;  Score 42;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              ||||||||
Db          1 SLLMWITQ 8
```

RESULT 17
AAB85304
ID AAB85304 standard; peptide; 9 AA.
XX
AC AAB85304;
XX
DT 17-SEP-2001 (first entry)
XX
DE HLA-A2 binding NY-ESO-1 peptide analogue.
XX
KW NY-ESO-1; human leukocyte antigen; HLA; lysis; cytolytic T cell; CTL;
KW HLA-A2; T-cell sorter; tumor; immune tetramer.
XX
OS Homo sapiens.
XX
PN WO200136453-A2.
XX
PD 25-MAY-2001.
XX
PF 08-NOV-2000; 2000WO-US042010.
XX
PR 15-NOV-1999; 99US-00440621.
PR 25-FEB-2000; 2000US-00514036.

PR 29-SEP-2000; 2000US-00676005.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
PA (UYOX-) UNIV OXFORD.
XX
PI Valmori D, Cerottini J, Romero P, Cerundolo V;
XX
DR WPI; 2001-451454/48.
XX
PT Novel isolated NY-ESO-1 nonapeptide useful for determining if a cell
PT presents human leukocyte antigen-A2 molecule on its surface, binds to
PT human leukocyte antigen molecules and provokes lysis by cytolytic T
PT cells.
XX
PS Example 19; Page 16; 38pp; English.
XX
CC The invention provides NY-ESO-1 peptide derivatives which bind to human
CC leukocyte antigen (HLA) molecules and provokes lysis by cytolytic T cells
CC (CTLs). The NY-ESO-1 nonapeptide is of formula SLLMWITQX, where X is an
CC amino acid having an uncharged polar side chain. The NY-ESO-1 peptide
CC derivatives are useful for determining if a cell presents an HLA-A2
CC molecule on its surface, by contacting a sample containing the cell with
CC the peptide or its derivative, and determining binding between them,
CC where the binding is indicative of HLA-A2 on the surface of the cell. The
CC NY-ESO-1 peptides and analogues are useful therapeutically, for
CC administration to a patient who is HLA-A2 positive and expresses NY-ESO-1
CC in connection with the pathology, as well as diagnostically, i.e. to
CC determine if HLA-A2 positive cells are present, or if relevant CTLs are
CC present. They are also useful for determining the presence of CTLs in a
CC sample. The peptides are useful as T-cell sorters, when incorporated into
CC immune tetramers. The present sequence represents a NY-ESO-1 peptide
CC analogue that can bind to HLA-A2 molecule
XX
SQ Sequence 9 AA;

Query Match 97.7%; Score 42; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLMWITQ 8
||||||||
Db 1 SLLMWITQ 8

RESULT 18
AAB85303
ID AAB85303 standard; peptide; 9 AA.
XX
AC AAB85303;
XX
DT 17-SEP-2001 (first entry)
XX
DE HLA-A2 binding NY-ESO-1 peptide.
XX
KW NY-ESO-1; human leukocyte antigen; HLA; lysis; cytolytic T cell; CTL;
KW HLA-A2; T-cell sorter; tumor; immune tetramer.
XX
OS Homo sapiens.
XX
PN WO200136453-A2.
XX
PD 25-MAY-2001.
XX
PF 08-NOV-2000; 2000WO-US042010.
XX
PR 15-NOV-1999; 99US-00440621.
PR 25-FEB-2000; 2000US-00514036.
PR 29-SEP-2000; 2000US-00676005.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
PA (UYOX-) UNIV OXFORD.
XX
PI Valmori D, Cerottini J, Romero P, Cerundolo V;
XX
DR WPI; 2001-451454/48.
XX
PT Novel isolated NY-ESO-1 nonapeptide useful for determining if a cell
PT presents human leukocyte antigen-A2 molecule on its surface, binds to
PT human leukocyte antigen molecules and provokes lysis by cytolytic T
PT cells.
XX
PS Claim 23; Page 16; 38pp; English.
XX
CC The invention provides NY-ESO-1 peptide derivatives which bind to human
CC leukocyte antigen (HLA) molecules and provokes lysis by cytolytic T cells
CC (CTLs). The NY-ESO-1 nonapeptide is of formula SLLMWITQX, where X is an
CC amino acid having an uncharged polar side chain. The NY-ESO-1 peptide
CC derivatives are useful for determining if a cell presents an HLA-A2
CC molecule on its surface, by contacting a sample containing the cell with
CC the peptide or its derivative, and determining binding between them,
CC where the binding is indicative of HLA-A2 on the surface of the cell. The
CC NY-ESO-1 peptides and analogues are useful therapeutically, for
CC administration to a patient who is HLA-A2 positive and expresses NY-ESO-1
CC in connection with the pathology, as well as diagnostically, i.e. to
CC determine if HLA-A2 positive cells are present, or if relevant CTLs are
CC present. They are also useful for determining the presence of CTLs in a
CC sample. The peptides are useful as T-cell sorters, when incorporated into
CC immune tetramers. The present sequence represents a NY-ESO-1 peptide that
CC can bind to HLA-A2 molecule
XX
SQ Sequence 9 AA;

Query Match 97.7%; Score 42; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLMWITQ 8
||||||||
Db 1 SLLMWITQ 8

RESULT 19
AAB85305
ID AAB85305 standard; peptide; 9 AA.
XX
AC AAB85305;
XX
DT 17-SEP-2001 (first entry)
XX
DE HLA-A2 binding NY-ESO-1 peptide analogue.
XX
KW NY-ESO-1; human leukocyte antigen; HLA; lysis; cytolytic T cell; CTL;
KW HLA-A2; T-cell sorter; tumor; immune tetramer.
XX
OS Homo sapiens.
XX
PN WO200136453-A2.
XX
PD 25-MAY-2001.
XX
PF 08-NOV-2000; 2000WO-US042010.
XX
PR 15-NOV-1999; 99US-00440621.
PR 25-FEB-2000; 2000US-00514036.
PR 29-SEP-2000; 2000US-00676005.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
PA (UYOX-) UNIV OXFORD.
XX
PI Valmori D, Cerottini J, Romero P, Cerundolo V;
XX
DR WPI; 2001-451454/48.
XX
PT Novel isolated NY-ESO-1 nonapeptide useful for determining if a cell
PT presents human leukocyte antigen-A2 molecule on its surface, binds to
PT human leukocyte antigen molecules and provokes lysis by cytolytic T

```
PT   cells.
XX
PS   Example 19; Page 16; 38pp; English.
XX
CC   The invention provides NY-ESO-1 peptide derivatives which bind to human
CC   leukocyte antigen (HLA) molecules and provokes lysis by cytolytic T cells
CC   (CTLs). The NY-ESO-1 nonapeptide is of formula SLLMWITQX, where X is an
CC   amino acid having an uncharged polar side chain. The NY-ESO-1 peptide
CC   derivatives are useful for determining if a cell presents an HLA-A2
CC   molecule on its surface, by contacting a sample containing the cell with
CC   the peptide or its derivative, and determining binding between them,
CC   where the binding is indicative of HLA-A2 on the surface of the cell. The
CC   NY-ESO-1 peptides and analogues are useful therapeutically, for
CC   administration to a patient who is HLA-A2 positive and expresses NY-ESO-1
CC   in connection with the pathology, as well as diagnostically, i.e. to
CC   determine if HLA-A2 positive cells are present, or if relevant CTLs are
CC   present. They are also useful for determining the presence of CTLs in a
CC   sample. The peptides are useful as T-cell sorters, when incorporated into
CC   immune tetramers. The present sequence represents a NY-ESO-1 peptide
CC   analogue that can bind to HLA-A2 molecule
XX
SQ   Sequence 9 AA;

  Query Match             97.7%;  Score 42;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              ||||||||
Db          1 SLLMWITQ 8


RESULT 20
AAB85308
ID   AAB85308 standard; peptide; 9 AA.
XX
AC   AAB85308;
XX
DT   17-SEP-2001  (first entry)
XX
DE   HLA-A2 binding NY-ESO-1 peptide #12.
XX
KW   NY-ESO-1; human leukocyte antigen; HLA; lysis; cytolytic T cell; CTL;
KW   HLA-A2; T-cell sorter; tumor; immune tetramer.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200136453-A2.
XX
PD   25-MAY-2001.
XX
PF   08-NOV-2000; 2000WO-US042010.
XX
PR   15-NOV-1999;   99US-00440621.
PR   25-FEB-2000; 2000US-00514036.
PR   29-SEP-2000; 2000US-00676005.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
PA   (UYOX-) UNIV OXFORD.
XX
PI   Valmori D,  Cerottini J,  Romero P,  Cerundolo V;
XX
DR   WPI; 2001-451454/48.
XX
PT   Novel isolated NY-ESO-1 nonapeptide useful for determining if a cell
PT   presents human leukocyte antigen-A2 molecule on its surface, binds to
PT   human leukocyte antigen molecules and provokes lysis by cytolytic T
PT   cells.
XX
PS   Example 14; Page 13; 38pp; English.
XX
CC   The invention provides NY-ESO-1 peptide derivatives which bind to human
CC   leukocyte antigen (HLA) molecules and provokes lysis by cytolytic T cells
CC   (CTLs). The NY-ESO-1 nonapeptide is of formula SLLMWITQX, where X is an
CC   amino acid having an uncharged polar side chain. The NY-ESO-1 peptide
CC   derivatives are useful for determining if a cell presents an HLA-A2
CC   molecule on its surface, by contacting a sample containing the cell with
CC   the peptide or its derivative, and determining binding between them,
CC   where the binding is indicative of HLA-A2 on the surface of the cell. The
CC   NY-ESO-1 peptides and analogues are useful therapeutically, for
CC   administration to a patient who is HLA-A2 positive and expresses NY-ESO-1
CC   in connection with the pathology, as well as diagnostically, i.e. to
CC   determine if HLA-A2 positive cells are present, or if relevant CTLs are
CC   present. They are also useful for determining the presence of CTLs in a
CC   sample. The peptides are useful as T-cell sorters, when incorporated into
CC   immune tetramers. The present sequence represents a NY-ESO-1 peptide that
CC   was tested for its binding capacity to a HLA-A2 molecule
XX
SQ   Sequence 9 AA;

  Query Match             97.7%;  Score 42;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+06;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              ||||||||
Db          1 SLLMWITQ 8


RESULT 21
AAB82017
ID   AAB82017 standard; peptide; 9 AA.
XX
AC   AAB82017;
XX
DT   12-JUN-2001  (first entry)
XX
DE   HLA- binding peptide derived from NY-ESO-1.
XX
KW   Multiple myeloma; tumour rejection antigen precursor; MAGE; BAGE; GAGE;
KW   LAGE; NY-ESO-1; PRAME; DAGE; human; HLA.
XX
OS   Homo sapiens.
XX
PN   US6210886-B1.
XX
PD   03-APR-2001.
XX
PF   30-OCT-1998;   98US-00183931.
XX
PR   04-FEB-1998;   98US-00018422.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Van Baren N,  Brasseur F,  Boon-Falleur T;
XX
DR   WPI; 2001-289628/30.
XX
PT   Detecting multiple myeloma in a patient, comprises contacting a nucleic
PT   acid containing sample taken from bone marrow or blood with a
PT   hybridization probe specific for a tumor rejection antigen precursor.
XX
PS   Example 3; Col 11; 16pp; English.
XX
CC   The present invention relates to a method for detecting multiple myeloma.
CC   The method comprises contacting a nucleic acid containing a sample taken
CC   from a bone marrow or blood of a patient, with a hybridisation probe
CC   specific for a tumour rejection antigen precursor. Tumour rejection
CC   antigen precursors used in the present invention are the MAGE family,
CC   BAGE, GAGE, LAGE, NY-ESO-1 and PRAME (previously referred to as DAGE).
CC   Expression of the tumour rejection antigen precursor indicates possible
CC   multiple myeloma in the patient. The method can also be used for
CC   monitoring the disease progress and course of therapeutic regime. The
CC   present sequence is a peptide derived from a tumour rejection antigen
CC   precursor, which was used in the method of the present invention
```

XX
SQ Sequence 9 AA;

Query Match 97.7%; Score 42; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLMWITQ 8
||||||||
Db 1 SLLMWITQ 8

RESULT 22
AAE06850
ID AAE06850 standard; peptide; 9 AA.
XX
AC AAE06850;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human NY-ESO-1 antigenic peptide #2.
XX
KW MAGE antigenic peptide; Human leukocyte antigen; HLA-B35; HLA-B44;
KW tumour cell; immunostimulant; antigen presentation; cancer; melanoma;
KW CD8+ cytotoxic T lymphocyte; colorectal; prostate; gastric carcinoma;
KW myeloma; brain tumour; sarcoma; seminoma; ovarian tumour; cytostatic;
KW gene therapy; tumour rejection antigen; TRA; human; NY-ESO-1; MHC;
KW major histocompatibility complex.
XX
OS Homo sapiens.
XX
PN WO200153833-A1.
XX
PD 26-JUL-2001.
XX
PF 19-JAN-2001; 2001WO-US002008.
XX
PR 20-JAN-2000; 2000US-0177242P.
PR 25-OCT-2000; 2000US-0243212P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Luiten R, Boon-Falleur T, Van Der Bruggen P, Stroobant V;
PI Demotte N, Schultz E;
XX
DR WPI; 2001-488724/53.
XX
PT Functional variants and isolated mimetics of a MAGE-A1 HLA-B35 or HLA-B44
PT binding peptide, or of a MAGE-A3 HLA-B35 binding peptide, used in
PT diagnosis and treatment of a disorder characterized by expression of MAGE
PT -A1 or -A3.
XX
PS Disclosure; Page 28; 103pp; English.
XX
CC The invention relates to functional variants and isolated mimetics of a
CC MAGE-A1 human leukocyte antigen (HLA)-B35 or HLA-B44 binding peptide, or
CC of a MAGE-A3 HLA-B35 binding peptide, identified by methods described in
CC the specification. MAGE genes encode tumour rejection antigens (TRAs)
CC presented to T lymphocytes by HLA-B35 and HLA-B44 molecules. The MAGE
CC antigenic peptide acts by binding to HLA molecules on tumour cells and
CC stimulating recognition of these cells and thus signalling them to the
CC immune system for destruction. The peptide when presented by HLA molecule
CC induces the activation and stimulation of CD8+ cytotoxic T lymphocytes.
CC The MAGE antigenic peptide is used to treat and diagnose disorders
CC characterised by expression of MAGE-A1 or -A3. Disorders include cancers
CC e.g melanomas, oesophageal, lung, head and neck, breast, colorectal,
CC prostate, renal, bladder, hepatocellular, papillary thyroid and gastric
CC carcinomas, myelomas, brain tumours, sarcomas, seminomas, and ovarian
CC tumours. The present sequence is human NY-ESO-1 tumour associated
CC antigenic peptide presented by major histocompatibility complex (MHC) HLA
CC -A2. The antigenic peptide is used in combination with peptides of the
CC invention for inducing an immune response
XX
SQ Sequence 9 AA;

Query Match 97.7%; Score 42; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLMWITQ 8
||||||||
Db 1 SLLMWITQ 8

RESULT 23
AAE26809
ID AAE26809 standard; peptide; 9 AA.
XX
AC AAE26809;
XX
DT 13-DEC-2002 (first entry)
XX
DE Human HLA-A2.1 restricted NY-ESO-1 peptide epitope #2.
XX
KW Human; cancer; breast cancer; ovarian cancer; melanoma; cell therapy;
KW epitope; human leucocyte antigen; HLA-A2.1.
XX
OS Homo sapiens.
XX
PN WO200265992-A2.
XX
PD 29-AUG-2002.
XX
PF 19-FEB-2002; 2002WO-US005748.
XX
PR 20-FEB-2001; 2001US-0270252P.
XX
PA (ORTH) ORTHO-MCNEIL PHARM INC.
XX
PI Degraw J, Moriarty A, Leturcq DJ, Jackson MR, Peterson PA;
PI Heiskala M;
XX
DR WPI; 2002-667033/71.
XX
PT Treating a subject with cancer comprises combining the CD+8 cells, which
PT are stimulated with non-naturally occurring antigen-presenting cell line,
PT with adherent blood monocytes and inoculating the subject with CD8+
PT suspension.
XX
PS Example 2; Page 94; 99pp; English.
XX
CC The invention relates to a method of treating a subject with cancer. The
CC method involves combining the CD+8 cells, which are stimulated with non
CC naturally occurring antigen-presenting cell (nnAPC) line, with adherent
CC blood monocytes and inoculating the subject with CD8+ suspension. The
CC method is useful for treating cancer e.g. ovarian cancer, breast cancer
CC and melanoma etc. It is also useful in cell therapy. The present sequence
CC is human leukocyte antigen A2 (HLA-A2).1 restricted peptide epitope used
CC to treat breast and ovarian cancer
XX
SQ Sequence 9 AA;

Query Match 97.7%; Score 42; DB 5; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLMWITQ 8
||||||||
Db 1 SLLMWITQ 8

RESULT 24
AAE26808
ID AAE26808 standard; peptide; 9 AA.
XX
AC AAE26808;

XX
DT 13-DEC-2002 (first entry)
XX
DE Human HLA-A2.1 restricted NY-ESO-1 peptide epitope #1.
XX
KW Human; cancer; breast cancer; ovarian cancer; melanoma; cell therapy;
KW epitope; human leucocyte antigen; HLA-A2.1.
XX
OS Homo sapiens.
XX
PN WO200265992-A2.
XX
PD 29-AUG-2002.
XX
PF 19-FEB-2002; 2002WO-US005748.
XX
PR 20-FEB-2001; 2001US-0270252P.
XX
PA (ORTH) ORTHO-MCNEIL PHARM INC.
XX
PI Degraw J, Moriarty A, Leturcq DJ, Jackson MR, Peterson PA;
PI Heiskala M;
XX
DR WPI; 2002-667033/71.
XX
PT Treating a subject with cancer comprises combining the CD+8 cells, which
PT are stimulated with non-naturally occurring antigen-presenting cell line,
PT with adherent blood monocytes and inoculating the subject with CD8+
PT suspension.
XX
PS Example 2; Page 94; 99pp; English.
XX
CC The invention relates to a method of treating a subject with cancer. The
CC method involves combining the CD+8 cells, which are stimulated with non
CC naturally occurring antigen-presenting cell (nnAPC) line, with adherent
CC blood monocytes and inoculating the subject with CD8+ suspension. The
CC method is useful for treating cancer e.g. ovarian cancer, breast cancer
CC and melanoma etc. It is also useful in cell therapy. The present sequence
CC is human leukocyte antigen A2 (HLA-A2).1 restricted peptide epitope used
CC to treat breast and ovarian cancer
XX
SQ Sequence 9 AA;

Query Match 97.7%; Score 42; DB 5; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLMWITQ 8
||||||||
Db 1 SLLMWITQ 8

RESULT 25
AAO21430
ID AAO21430 standard; peptide; 9 AA.
XX
AC AAO21430;
XX
DT 06-AUG-2002 (first entry)
XX
DE HLA-A2 restricted peptide sequence.
XX
KW Immunostimulant; human leukocyte antigen; HLA-Cw3; HLA-Cw6; cytolytic;
KW proliferation; T cell; HLA-CW3/HLA-CW6; HLA-A2.
XX
OS Homo sapiens.
XX
PN WO200226778-A2.
XX
PD 04-APR-2002.
XX
PF 24-SEP-2001; 2001WO-US029920.
XX
PR 26-SEP-2000; 2000US-00670456.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Gnjatic S, Old LJ, Nagata Y, Jager E, Chen Y, Knuth A;
XX
DR WPI; 2002-435193/46.
XX
PT Novel isolated human peptide that binds to human leukocyte antigen-Cw3 or
PT HLA-Cw6, useful for stimulating proliferation of cytolytic T cells.
XX
PS Example 3; Page 9; 21pp; English.
XX
CC The invention relates to an isolated peptide which binds to a human
CC leukocyte antigen (HLA)-Cw3 molecule or binds to a HLA-Cw6 molecule. The
CC isolated peptide provokes proliferation of T cells specific to a complex
CC of the isolated peptide and HLA-Cw3, or the isolated peptide and HLA-Cw6.
CC The isolated peptide is useful for stimulating proliferation of a
CC cytolytic T cell response, by contacting a T cell containing sample with
CC a cell which presents a complex of HLA-CW3/HLA-CW6 and the isolated
CC peptide of the invention on its surface. This sequence represents an HLA-
CC A2 restricted peptide sequence relating to the invention
XX
SQ Sequence 9 AA;

Query Match 97.7%; Score 42; DB 5; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLMWITQ 8
||||||||
Db 1 SLLMWITQ 8

Search completed: August 22, 2004, 10:59:56
Job time : 123 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 22, 2004, 10:54:38 ; Search time 113 Seconds
(without alignments)
25.130 Million cell updates/sec

Title: US-10-706-475-10
Perfect score: 43
Sequence: 1 SLLMWITQX 9

Scoring table: BLOSUM62XX
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	42	97.7	180	4	Q9Y479	Q9y479 homo sapien
2	36	83.7	198	16	Q8UCP4	Q8ucp4 agrobacteri
3	36	83.7	401	11	Q8C0T0	Q8c0t0 mus musculu
4	35	81.4	861	16	O06944	O06944 synechocyst
5	34	79.1	202	5	Q95YB6	Q95yb6 caenorhabdi
6	34	79.1	251	16	Q8KE76	Q8ke76 chlorobium
7	34	79.1	299	10	Q9ZVK1	Q9zvk1 arabidopsis
8	34	79.1	530	4	Q9C0G1	Q9c0g1 homo sapien
9	34	79.1	620	17	Q96XC8	Q96xc8 sulfolobus
10	34	79.1	1137	5	Q9VXF4	Q9vxf4 drosophila
11	34	79.1	1137	5	Q8T9H7	Q8t9h7 drosophila
12	33	76.7	122	8	Q7YAI1	Q7yai1 percus guir
13	33	76.7	227	16	Q7WQK1	Q7wqk1 bordetella
14	33	76.7	227	16	Q7WCJ5	Q7wcj5 bordetella
15	33	76.7	227	16	Q7VZW4	Q7vzw4 bordetella
16	33	76.7	362	16	Q7VF34	Q7vf34 helicobacte
17	33	76.7	479	17	Q9HSK2	Q9hsk2 halobacteri
18	33	76.7	571	10	Q7XE56	Q7xe56 oryza sativ
19	33	76.7	683	16	Q82WC4	Q82wc4 nitrosomona
20	33	76.7	713	10	Q9FGS9	Q9fgs9 arabidopsis
21	33	76.7	728	10	Q8VY13	Q8vy13 arabidopsis
22	32	74.4	121	5	Q8IEI7	Q8iei7 plasmodium
23	32	74.4	127	8	Q7YBD6	Q7ybd6 tapinella s
24	32	74.4	171	4	Q9Y2R7	Q9y2r7 homo sapien
25	32	74.4	229	4	Q96FF6	Q96ff6 homo sapien
26	32	74.4	240	16	Q89YL2	Q89yl2 bacteroides
27	32	74.4	251	2	Q93GH0	Q93gh0 bacillus su
28	32	74.4	251	2	O52853	O52853 bacillus su
29	32	74.4	278	16	Q9JZQ7	Q9jzq7 neisseria m
30	32	74.4	424	4	Q8IXA2	Q8ixa2 homo sapien
31	32	74.4	454	4	Q8N1C5	Q8n1c5 homo sapien
32	32	74.4	454	11	Q80UI5	Q80ui5 mus musculu
33	32	74.4	458	16	Q8FEC2	Q8fec2 escherichia
34	32	74.4	721	10	Q9LTA0	Q9lta0 arabidopsis
35	32	74.4	722	13	Q8UUU2	Q8uuu2 xenopus lae
36	32	74.4	724	11	P70304	P70304 mus musculu
37	32	74.4	724	11	Q8K3B3	Q8k3b3 mus musculu
38	32	74.4	738	10	Q8RWS6	Q8rws6 arabidopsis
39	32	74.4	739	10	Q9LT99	Q9lt99 arabidopsis
40	32	74.4	753	10	Q7XUW3	Q7xuw3 oryza sativ
41	32	74.4	872	16	Q9RIU6	Q9riu6 streptomyce
42	32	74.4	999	16	Q8YMR7	Q8ymr7 anabaena sp
43	32	74.4	1076	10	Q9SHX8	Q9shx8 arabidopsis
44	31	72.1	101	10	Q9LKL7	Q9lkl7 vitis berla
45	31	72.1	104	4	Q8TAY4	Q8tay4 homo sapien
46	31	72.1	122	8	Q8HH81	Q8hh81 diglymma cl
47	31	72.1	124	16	Q82Q49	Q82q49 streptomyce
48	31	72.1	130	8	Q8HH80	Q8hh80 diglymma cl
49	31	72.1	132	8	Q8HH79	Q8hh79 diglymma cl
50	31	72.1	134	16	Q8F6X7	Q8f6x7 leptospira

ALIGNMENTS

RESULT 1
Q9Y479
ID Q9Y479 PRELIMINARY; PRT; 180 AA.
AC Q9Y479;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE LAGE-1S protein (Cancer/testis antigen 2).
GN LAGE1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=99325550; PubMed=10399963;
RA Aarnoudse C.A., Van den Doel P.B., Heemskerk B., Schrier P.I.;
RT "Interleukin-2-induced, melanoma-specific T cells recognize CAMEL, an
RT unexpected translation product of LAGE-1.";
RL Int. J. Cancer 82:442-448(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Aradhya S., Bardaro T., Galgoczy P., Yamagata T., Esposito T.,
RA Patlan H., Ciccodicola A., Kenwrick S., Platzer M., D'Urso M.,
RA Nelson D.L.;
RT "Multiple pathogenic and benign genomic rearrangements occur at a 35-
RT kb duplication involving the NEMO and the LAGE2 genes.";
RL Hum. Mol. Genet. 0:0-0(2001).
DR EMBL; AJ012834; CAA10194.1; -.
DR EMBL; AF277315; AAL27015.1; -.
SQ SEQUENCE 180 AA; 18236 MW; 9077FAF953543A25 CRC64;

Query Match 97.7%; Score 42; DB 4; Length 180;

```
  Best Local Similarity   100.0%;  Pred. No. 3.9;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              ||||||||
Db        157 SLLMWITQ 164


RESULT 2
Q8UCP4
ID   Q8UCP4      PRELIMINARY;      PRT;   198 AA.
AC   Q8UCP4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein Atu2439.
GN   ATU2439 OR AGR_C_4424.
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX   NCBI_TaxID=176299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608550; PubMed=11743193;
RA   Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA   Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA   Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA   Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA   Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA   Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA   Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA   Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA   Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA   Nester E.W.;
RT   "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT   C58.";
RL   Science 294:2317-2323(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608551; PubMed=11743194;
RA   Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA   Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA   Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA   Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA   Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA   Cielo C., Slater S.;
RT   "Genome sequence of the plant pathogen and biotechnology agent
RT   Agrobacterium tumefaciens C58.";
RL   Science 294:2323-2328(2001).
DR   EMBL; AE009191; AAL43427.1; ALT_INIT.
DR   EMBL; AE008157; AAK88176.1; -.
DR   InterPro; IPR008325; UCP033924.
DR   PIRSF; PIRSF033924; UCP033924; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   198 AA;  21154 MW;  FA40C6E548F0562F CRC64;

  Query Match             83.7%;  Score 36;  DB 16;  Length 198;
  Best Local Similarity   85.7%;  Pred. No. 52;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWIT 7
              |||||:|
Db         25 SLLMWVT 31


RESULT 3
Q8C0T0
ID   Q8C0T0      PRELIMINARY;      PRT;   401 AA.
AC   Q8C0T0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical type-1 copper.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK029920; BAC26677.1; -.
DR   GO; GO:0005507; F:copper ion binding; IEA.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR000923; BlueCu_1.
DR   PROSITE; PS00196; COPPER_BLUE; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   401 AA;  45562 MW;  F34F237653D6FCF9 CRC64;

  Query Match             83.7%;  Score 36;  DB 11;  Length 401;
  Best Local Similarity   85.7%;  Pred. No. 1e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWIT 7
              |||||:|
Db        316 SLLMWVT 322


RESULT 4
O06944
ID   O06944      PRELIMINARY;      PRT;   861 AA.
AC   O06944; Q55995;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein sll0737.
GN   SLL0737.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96127529; PubMed=8590279;
RA   Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA   Sugiura M., Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT   region from map positions 64% to 92% of the genome.";
RL   DNA Res. 2:153-166(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA   Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA   Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
DR   EMBL; D64005; BAA10778.1; -.
DR   PIR; S77086; S77086.
DR   InterPro; IPR008941; TPR-like.
DR   InterPro; IPR007016; Wzy_C.
DR   Pfam; PF04932; Wzy_C; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   861 AA;  96682 MW;  A064B98C2D9B6C59 CRC64;

  Query Match             81.4%;  Score 35;  DB 16;  Length 861;
```

```
  Best Local Similarity   75.0%;  Pred. No. 3.1e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              |||:| ||
Db        157 SLLLWFTQ 164


RESULT 5
Q95YB6
ID   Q95YB6      PRELIMINARY;      PRT;   202 AA.
AC   Q95YB6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   F28A10.8.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Miller N., Wamsley P., Gibson A.;
RT   "The sequence of C. elegans cosmid F28A10.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006632; AAK85474.1; -.
DR   WormPep; F28A10.8; CE19413.
DR   InterPro; IPR005515; VOMI.
DR   Pfam; PF03762; VOMI; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   202 AA;  22172 MW;  129FDF141832D165 CRC64;

  Query Match             79.1%;  Score 34;  DB 5;  Length 202;
  Best Local Similarity   85.7%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWIT 7
              |||:|||
Db         10 SLLLWIT 16


RESULT 6
Q8KE76
ID   Q8KE76      PRELIMINARY;      PRT;   251 AA.
AC   Q8KE76;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein CT0814.
GN   CT0814.
OS   Chlorobium tepidum.
OC   Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC   Chlorobium.
OX   NCBI_TaxID=1097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TLS / ATCC 49652 / DSM 12025;
RX   MEDLINE=22103685; PubMed=12093901;
RA   Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA   Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
RA   Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA   Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA   Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
RA   Vamathevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
RA   Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
RT   "The complete genome sequence of Chlorobium tepidum TLS, a
RT   photosynthetic, anaerobic, green-sulfur bacterium.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
DR   EMBL; AE012849; AAM72050.1; -.
DR   TIGR; CT0814; -.
DR   GO; GO:0016491; F:oxidoreductase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR000415; Nitroreductase.
DR   Pfam; PF00881; Nitroreductase; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   251 AA;  27932 MW;  F7FDA5ED4FCF7985 CRC64;

  Query Match             79.1%;  Score 34;  DB 16;  Length 251;
  Best Local Similarity   75.0%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              |||:| ||
Db         87 SLLLWATQ 94


RESULT 7
Q9ZVK1
ID   Q9ZVK1      PRELIMINARY;      PRT;   299 AA.
AC   Q9ZVK1;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative endoxyloglucan glycosyltransferase.
GN   AT2G14620.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Banh J., Chang C.H., Chang E., Dale J.M., Goldsmith A.D.,
RA   Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
RA   Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA   Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA   Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M.,
RA   Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA   Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA   Theologis A.;
RT   "Full Length cDNA of gene At2g14620 (GI:15225976).";
```

```
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA   Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
RA   Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA   Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA   Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA   Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
RA   Davis R.W., Ecker J.R., Theologis A.;
RT   "Arabidopsis Open Reading Frame (ORF) Clones.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC005398; AAC69380.1; -.
DR   EMBL; AY070415; AAL49911.1; -.
DR   EMBL; AY096596; AAM20246.1; -.
DR   PIR; D84519; D84519.
DR   HSSP; P23904; 1AJK.
DR   GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000757; Glyco_hydro_16.
DR   Pfam; PF00722; Glyco_hydro_16; 1.
KW   Transferase.
SQ   SEQUENCE   299 AA;  34687 MW;  C8A688EFB7E910A5 CRC64;

  Query Match             79.1%;  Score 34;  DB 10;  Length 299;
  Best Local Similarity   62.5%;  Pred. No. 1.8e+02;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              |||:|::|
Db         21 SLLLWVSQ 28


RESULT 8
Q9C0G1
ID   Q9C0G1      PRELIMINARY;      PRT;   530 AA.
AC   Q9C0G1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein KIAA1702 (Fragment).
GN   KIAA1702.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21082932; PubMed=11214970;
RA   Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIX.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 7:347-355(2000).
DR   EMBL; AB051489; BAB21793.1; -.
DR   GO; GO:0005351; F:sugar porter activity; IEA.
DR   GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. . .; IEA.
DR   InterPro; IPR002114; HPr_SerP_S.
DR   PROSITE; PS00589; PTS_HPR_SER; 1.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   530 AA;  60257 MW;  47638AA273036481 CRC64;

  Query Match             79.1%;  Score 34;  DB 4;  Length 530;
  Best Local Similarity   85.7%;  Pred. No. 3e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LLMWITQ 8
              | |||||
Db         27 LFMWITQ 33


RESULT 9
Q96XC8
ID   Q96XC8      PRELIMINARY;      PRT;   620 AA.
AC   Q96XC8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein ST2587.
GN   ST2587.
OS   Sulfolobus tokodaii.
OC   Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC   Sulfolobus.
OX   NCBI_TaxID=111955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JCM 10545 / 7;
RX   MEDLINE=21456156; PubMed=11572479;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA   Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA   Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA   Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA   Oshima T., Kikuchi H.;
RT   "Complete genome sequence of an aerobic thermoacidophilic
RT   Crenarchaeon, Sulfolobus tokodaii strain7.";
RL   DNA Res. 8:123-140(2001).
DR   EMBL; AP000990; BAB67700.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   620 AA;  68673 MW;  FE398E1860560918 CRC64;

  Query Match             79.1%;  Score 34;  DB 17;  Length 620;
  Best Local Similarity   85.7%;  Pred. No. 3.5e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWIT 7
              |||:|||
Db        328 SLLLWIT 334


RESULT 10
Q9VXF4
ID   Q9VXF4      PRELIMINARY;      PRT;  1137 AA.
AC   Q9VXF4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CG9902 protein.
GN   CG9902.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
```

```
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003502; AAF48620.1; -.
DR   FlyBase; FBgn0030757; CG9902.
SQ   SEQUENCE   1137 AA;  129332 MW;  4DE52C0D7E792B21 CRC64;

  Query Match           79.1%;  Score 34;  DB 5;  Length 1137;
  Best Local Similarity   75.0%;  Pred. No. 6e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              ||||| :|
Db         88 SLLMWFSQ 95


RESULT 11
Q8T9H7
ID   Q8T9H7      PRELIMINARY;      PRT;  1137 AA.
AC   Q8T9H7;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   GM14421p.
GN   CG9902.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY069289; AAL39434.1; -.
DR   FlyBase; FBgn0030757; CG9902.
SQ   SEQUENCE   1137 AA;  129366 MW;  FF16FBA33647D3B6 CRC64;

  Query Match           79.1%;  Score 34;  DB 5;  Length 1137;
  Best Local Similarity   75.0%;  Pred. No. 6e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              ||||| :|
Db         88 SLLMWFSQ 95


RESULT 12
Q7YAI1
ID   Q7YAI1      PRELIMINARY;      PRT;  122 AA.
AC   Q7YAI1;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NADH dehydrogenase subunit 1 (Fragment).
GN   ND1.
OS   Percus guiraoi.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Coleoptera; Adephaga; Carabidae;
OC   Carabidae Conjunctae; Percus.
OX   NCBI_TaxID=239177;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Brueckner M.;
RT   "The phylogeny of the subgenus Pseudopercus (Coleoptera, Carabidae):
RT   Conflicts between mitochondrial and nuclear DNA sequences.";
RL   Mitt. Dtsch. Ges. Allg. Angew. Entomol. 0:0-0(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Brueckner M.;
RT   "Phylogeny and biogeography of the genus Percus Bonelli 1810
RT   (Coleoptera, Carabidae): A molecular systematic analysis.";
RL   Thesis (2002), University of Bremen, Germany.
DR   EMBL; AY334343; AAQ01647.1; -.
KW   Mitochondrion.
FT   NON_TER     122    122
SQ   SEQUENCE   122 AA;  14051 MW;  7B05ED6C97CB06B0 CRC64;

  Query Match           76.7%;  Score 33;  DB 8;  Length 122;
  Best Local Similarity   71.4%;  Pred. No. 1.2e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWIT 7
              |||:|:|
Db         86 SLLLWVT 92


RESULT 13
Q7WQK1
ID   Q7WQK1      PRELIMINARY;      PRT;  227 AA.
AC   Q7WQK1;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Glutamate/aspartate transport system permease protein.
GN   GLTK OR BB0331.
OS   Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Bordetella.
OX   NCBI_TaxID=518;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RB50 / ATCC BAA-588;
RX   MEDLINE=22827954; PubMed=12910271;
RA   Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA   Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA   Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA   Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA   Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA   Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA   Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA   Rabbinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA   Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA   Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT   "Comparative analysis of the genome sequences of Bordetella pertussis,
RT   Bordetella parapertussis and Bordetella bronchiseptica.";
RL   Nat. Genet. 35:32-40(2003).
DR   EMBL; BX640437; CAE30829.1; -.
KW   Complete proteome.
SQ   SEQUENCE   227 AA;  24798 MW;  4A5A84A73EEFB1E7 CRC64;
```

```
  Query Match             76.7%;  Score 33;  DB 16;  Length 227;
  Best Local Similarity   85.7%;  Pred. No. 2.1e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LLMWITQ 8
              || ||||
Db         83 LLQWITQ 89


RESULT 14
Q7WCJ5
ID   Q7WCJ5      PRELIMINARY;      PRT;   227 AA.
AC   Q7WCJ5;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Glutamate/aspartate transport system permease protein.
GN   GLTK OR BPP0328.
OS   Bordetella parapertussis.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Bordetella.
OX   NCBI_TaxID=519;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=12822 / ATCC BAA-587;
RX   MEDLINE=22827954; PubMed=12910271;
RA   Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA   Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA   Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA   Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA   Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA   Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA   Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA   Rabbinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA   Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA   Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT   "Comparative analysis of the genome sequences of Bordetella pertussis,
RT   Bordetella parapertussis and Bordetella bronchiseptica.";
RL   Nat. Genet. 35:32-40(2003).
DR   EMBL; BX640424; CAE35912.1; -.
KW   Complete proteome.
SQ   SEQUENCE   227 AA;  24814 MW;  AF0FC4A8B0D8D3CA CRC64;

  Query Match             76.7%;  Score 33;  DB 16;  Length 227;
  Best Local Similarity   85.7%;  Pred. No. 2.1e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LLMWITQ 8
              || ||||
Db         83 LLQWITQ 89


RESULT 15
Q7VZW4
ID   Q7VZW4      PRELIMINARY;      PRT;   227 AA.
AC   Q7VZW4;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Glutamate/aspartate transport system permease protein.
GN   GLTK OR BP0767.
OS   Bordetella pertussis.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Bordetella.
OX   NCBI_TaxID=520;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RX   MEDLINE=22827954; PubMed=12910271;
RA   Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA   Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA   Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA   Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA   Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA   Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA   Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA   Rabbinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA   Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA   Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT   "Comparative analysis of the genome sequences of Bordetella pertussis,
RT   Bordetella parapertussis and Bordetella bronchiseptica.";
RL   Nat. Genet. 35:32-40(2003).
DR   EMBL; BX640413; CAE41073.1; -.
KW   Complete proteome.
SQ   SEQUENCE   227 AA;  24798 MW;  4A5A84A73EEFB1E7 CRC64;

  Query Match             76.7%;  Score 33;  DB 16;  Length 227;
  Best Local Similarity   85.7%;  Pred. No. 2.1e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LLMWITQ 8
              || ||||
Db         83 LLQWITQ 89


RESULT 16
Q7VF34
ID   Q7VF34      PRELIMINARY;      PRT;   362 AA.
AC   Q7VF34;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Integrase/recombinase XerD.
GN   XERD OR HH1845.
OS   Helicobacter hepaticus.
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=32025;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 51449 / 3B1;
RX   MEDLINE=22709201; PubMed=12810954;
RA   Suerbaum S., Josenhans C., Sterzenbach T., Drescher B., Brandt P.,
RA   Bell M., Droege M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,
RA   Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,
RA   Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.;
RT   "The complete genome sequence of the carcinogenic bacterium
RT   Helicobacter hepaticus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906(2003).
DR   EMBL; AE017149; AAP78442.1; -.
KW   Complete proteome.
SQ   SEQUENCE   362 AA;  41280 MW;  581A08F96F841C0D CRC64;

  Query Match             76.7%;  Score 33;  DB 16;  Length 362;
  Best Local Similarity   75.0%;  Pred. No. 3.2e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              ||| |||:
Db         14 SLLFWITR 21


RESULT 17
Q9HSK2
ID   Q9HSK2      PRELIMINARY;      PRT;   479 AA.
AC   Q9HSK2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Vng0189c.
GN   VNG0189C.
OS   Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR   EMBL; AE004984; AAG18802.1; -.
DR   PIR; F84179; F84179.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   InterPro; IPR006674; HD.
DR   InterPro; IPR003607; Met_phsphohydro.
DR   Pfam; PF01966; HD; 1.
DR   SMART; SM00471; HDc; 1.
KW   Complete proteome.
SQ   SEQUENCE   479 AA;  53083 MW;  9081FC3F14BE8C1B CRC64;

  Query Match             76.7%;  Score 33;  DB 17;  Length 479;
  Best Local Similarity   62.5%;  Pred. No. 4.2e+02;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              :| ||:|| .
Db        263 ALYMWVTQ 270


RESULT 18
Q7XE56
ID   Q7XE56      PRELIMINARY;      PRT;   571 AA.
AC   Q7XE56;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   OSJNBB0040I11.1.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   The Rice Chromosome 10 Sequencing Consortium;
RT   "In-depth view of structure, activity, and evolution of rice
RT   chromosome 10.";
RL   Science 300:1566-1569(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE017098; AAP53934.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   571 AA;  63008 MW;  401866DAAC2A8F5F CRC64;

  Query Match             76.7%;  Score 33;  DB 10;  Length 571;
  Best Local Similarity   75.0%;  Pred. No. 4.9e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              :|| ||||
Db        403 TLLTWITQ 410


RESULT 19
Q82WC4
ID   Q82WC4      PRELIMINARY;      PRT;   683 AA.
AC   Q82WC4;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cytochrome c-type biogenesis protein (CcmF).
GN   CCMF OR NE0768.
OS   Nitrosomonas europaea.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC   Nitrosomonadaceae; Nitrosomonas.
OX   NCBI_TaxID=915;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19718 / IFO 14298;
RX   MEDLINE=22586410; PubMed=12700255;
RA   Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
RA   Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
RA   Arciero D.M., Hommes N.G., Whittaker M.M., Arp D.J.;
RT   "Complete genome sequence of the ammonia-oxidizing bacterium and
RT   obligate chemolithoautotroph Nitrosomonas europaea.";
RL   J. Bacteriol. 185:2759-2773(2003).
DR   EMBL; BX321858; CAD84679.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0015232; F:heme transporter activity; IEA.
DR   GO; GO:0017004; P:cytochrome biogenesis; IEA.
DR   GO; GO:0008535; P:cytochrome c oxidase biogenesis; IEA.
DR   GO; GO:0015886; P:heme transport; IEA.
DR   InterPro; IPR002541; CytC_asm.
DR   InterPro; IPR003568; Cytc_biog_CcmF.
DR   InterPro; IPR003567; Cyt_c_biog.
DR   Pfam; PF01578; CytC_asm; 1.
DR   PRINTS; PR01410; CCBIOGENESIS.
DR   TIGRFAMs; TIGR00353; nrfE; 1.
KW   Complete proteome.
SQ   SEQUENCE   683 AA;  74061 MW;  6A0624E452D77FC9 CRC64;

  Query Match             76.7%;  Score 33;  DB 16;  Length 683;
  Best Local Similarity   71.4%;  Pred. No. 5.8e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWIT 7
              |||:|:|
Db         95 SLLLWVT 101


RESULT 20
Q9FGS9
ID   Q9FGS9      PRELIMINARY;      PRT;   713 AA.
AC   Q9FGS9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   FRO1 and FRO2-like protein.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Columbia;
RX   MEDLINE=20181125; PubMed=10718197;
RA   Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT   features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT   clones.";
RL   DNA Res. 7:31-63(2000).
DR   EMBL; AB024031; BAB09387.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0016491; F:oxidoreductase activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
```

```
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR000515; BPD_transp.
DR   InterPro; IPR002916; Ferric_reduct.
DR   InterPro; IPR000778; GP91PhoX.
DR   Pfam; PF01794; Ferric_reduct; 1.
DR   PRINTS; PR00466; GP91PHOX.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; 1.
SQ   SEQUENCE   713 AA;  81511 MW;  A2455CCFAEBB8310 CRC64;

  Query Match           76.7%;  Score 33;  DB 10;  Length 713;
  Best Local Similarity  100.0%;  Pred. No. 6e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LLMWIT 7
              ||||||
Db        248 LLMWIT 253


RESULT 21
Q8VY13
ID   Q8VY13      PRELIMINARY;      PRT;   728 AA.
AC   Q8VY13;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Putative FRO1 and FRO2.
GN   AT5G50160.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA   Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
RA   Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA   Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA   Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA   Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
RA   Davis R.W., Ecker J.R., Theologis A.;
RT   "Full Length cDNA of gene At5g50160 (GI:15240624).";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
RA   Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,
RA   Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
RA   Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,
RA   Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
RA   Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,
RA   Theologis A.;
RT   "Arabidopsis Open Reading Frame (ORF) Clones.";
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY074287; AAL66984.1; -.
DR   EMBL; BT003007; AAO22815.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0016491; F:oxidoreductase activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR000515; BPD_transp.
DR   InterPro; IPR002916; Ferric_reduct.
DR   InterPro; IPR000778; GP91PhoX.
DR   Pfam; PF01794; Ferric_reduct; 1.
DR   PRINTS; PR00466; GP91PHOX.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; 1.
SQ   SEQUENCE   728 AA;  83230 MW;  A1E5BFD18FB1D2B0 CRC64;

  Query Match           76.7%;  Score 33;  DB 10;  Length 728;
  Best Local Similarity  100.0%;  Pred. No. 6.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LLMWIT 7
              ||||||
Db        248 LLMWIT 253


RESULT 22
Q8IEI7
ID   Q8IEI7      PRELIMINARY;      PRT;   121 AA.
AC   Q8IEI7;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   MAL13P1.62.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,
RA   Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
RA   Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL844509; CAD52270.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   121 AA;  15146 MW;  C99C76BE87B2F5D6 CRC64;

  Query Match           74.4%;  Score 32;  DB 5;  Length 121;
  Best Local Similarity  100.0%;  Pred. No. 1.8e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWI 6
              ||||||
Db         17 SLLMWI 22


RESULT 23
Q7YBD6
ID   Q7YBD6      PRELIMINARY;      PRT;   127 AA.
AC   Q7YBD6;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cytochrome oxidase subunit I (Fragment).
GN   COI.
OS   Tapinella sp. KPJ-2003.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Paraneoptera; Psocoptera; Troctomorpha; Pachytroctidae;
OC   Tapinella.
OX   NCBI_TaxID=239299;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Johnson K.P., Mockford E.L.;
RT   "Molecular systematics of Psocomorpha (Psocoptera).";
RL   Syst. Entomol. 28:409-416(2003).
DR   EMBL; AY275295; AAP97113.1; -.
KW   Mitochondrion.
FT   NON_TER       1      1
FT   NON_TER     127    127
SQ   SEQUENCE   127 AA;  13949 MW;  9EF19D85424BF551 CRC64;

  Query Match           74.4%;  Score 32;  DB 8;  Length 127;
  Best Local Similarity  100.0%;  Pred. No. 1.9e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWI 6
              ||||||
Db         96 SLLMWI 101


RESULT 24
```

```
Q9Y2R7
ID   Q9Y2R7      PRELIMINARY;      PRT;   171 AA.
AC   Q9Y2R7;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   HSPC013 (Apoptosis related protein APR-3).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20499367; PubMed=11042152;
RA   Zhang Q.H., Ye M., Wu X.Y., Ren S.X., Zhao M., Zhao C.J., Fu G.,
RA   Shen Y., Fan H.Y., Lu G., Zhong M., Xu X.R., Han Z.G., Zhang J.W.,
RA   Tao J., Huang Q.H., Zhou J., Hu G.X., Gu J., Chen S.J., Chen Z.;
RT   "Cloning and functional analysis of cDNAs with open reading frames for
RT   300 previously undefined genes expressed in CD34+ hematopoietic
RT   stem/progenitor cells.";
RL   Genome Res. 10:1546-1560(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF077037; AAD27770.1; -.
DR   EMBL; BC035850; AAH35850.1; -.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
KW   EGF-like domain.
SQ   SEQUENCE   171 AA;  18597 MW;  91A42CD2B2CB0883 CRC64;

  Query Match             74.4%;  Score 32;  DB 4;  Length 171;
  Best Local Similarity   62.5%;  Pred. No. 2.5e+02;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              |:|:| ||
Db        157 SILLWATQ 164


RESULT 25
Q96FF6
ID   Q96FF6      PRELIMINARY;      PRT;   229 AA.
AC   Q96FF6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC011006; AAH11006.1; -.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
KW   Hypothetical protein; EGF-like domain.
SQ   SEQUENCE   229 AA;  24688 MW;  F2C63F934A47ED33 CRC64;

  Query Match             74.4%;  Score 32;  DB 4;  Length 229;
  Best Local Similarity   62.5%;  Pred. No. 3.2e+02;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              |:|:| ||
Db        215 SILLWATQ 222


Search completed: August 22, 2004, 11:02:24
Job time : 114 secs
```

This Page Blank (uspto)

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 22, 2004, 10:47:33 ; Search time 23 Seconds
(without alignments)
20.375 Million cell updates/sec

Title: US-10-706-475-10
Perfect score: 43
Sequence: 1 SLLMWITQX 9

Scoring table: BLOSUM62XX
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	42	97.7	180	1	CTG1_HUMAN	P78358 homo sapien
2	34	79.1	122	1	YM10_YEAST	Q03880 saccharomyc
3	34	79.1	299	1	XT10_ARATH	Q9zvk1 arabidopsis
4	32	74.4	145	1	YD54_METJA	Q58749 methanococc
5	32	74.4	724	1	P85A_BOVIN	P23727 bos taurus
6	32	74.4	724	1	P85A_HUMAN	P27986 homo sapien
7	32	74.4	724	1	P85A_MOUSE	P26450 mus musculu
8	32	74.4	724	1	P85A_RAT	Q63787 rattus norv
9	32	74.4	2331	1	RRPL_MABVM	P31352 marburg vir
10	32	74.4	2331	1	RRPL_MABVP	P35262 marburg vir
11	31	72.1	213	1	NUKM_SOLTU	Q43844 solanum tub
12	31	72.1	356	1	CKR8_MACMU	O97665 macaca mula
13	31	72.1	437	1	PFTB_BOVIN	P49355 bos taurus
14	31	72.1	437	1	PFTB_HUMAN	P49356 homo sapien
15	31	72.1	437	1	PFTB_RAT	Q02293 rattus norv
16	31	72.1	6885	1	SNE2_HUMAN	Q8wxh0 homo sapien
17	31	72.1	8797	1	SNE1_HUMAN	Q8nf91 homo sapien
18	30	69.8	83	1	VE5_HPV16	P06927 human papil
19	30	69.8	121	1	KV40_HUMAN	P06312 homo sapien
20	30	69.8	133	1	KV4B_HUMAN	P06313 homo sapien
21	30	69.8	134	1	KV4C_HUMAN	P06314 homo sapien
22	30	69.8	351	1	HRPX_PLALO	P04929 plasmodium
23	30	69.8	406	1	HOFC_HAEIN	P44621 haemophilus
24	30	69.8	675	1	RGS9_MOUSE	O54828 mus musculu
25	30	69.8	676	1	KALM_CHICK	P33005 gallus gall
26	30	69.8	702	1	HPPA_RHORU	O68460 rhodospiril
27	30	69.8	706	1	HPPA_BRAJA	Q89k83 bradyrhizob
28	30	69.8	806	1	RIR1_HSV7J	P50641 human herpe
29	30	69.8	2179	1	POLG_EC23W	O73556 e genome po
30	30	69.8	2180	1	POLG_EC22H	Q66578 e genome po
31	30	69.8	2188	1	POLG_EC23C	Q9yid8 e genome po
32	29	67.4	68	1	NP31_MOUSE	Q07475 mus musculu
33	29	67.4	98	1	NULM_POLOR	Q95916 polypterus
34	29	67.4	117	1	V117_ASFL5	P26702 african swi
35	29	67.4	184	1	21KD_ZYMMO	P15256 zymomonas m
36	29	67.4	207	1	TIM1_CANFA	P81546 canis famil
37	29	67.4	234	1	SNG4_HUMAN	O95473 homo sapien
38	29	67.4	240	1	YCXN_PORPU	P51364 porphyra pu
39	29	67.4	284	1	XT13_ARATH	Q9fk18 arabidopsis
40	29	67.4	305	1	NU1M_APILI	P34847 apis mellif
41	29	67.4	326	1	NU37_HUMAN	Q8nfh4 homo sapien
42	29	67.4	326	1	NU37_MOUSE	Q9cwu9 mus musculu
43	29	67.4	328	1	O4AG_HUMAN	Q8nh70 homo sapien
44	29	67.4	402	1	OPS4_CANAL	P46596 candida alb
45	29	67.4	402	1	YYCB_BACSU	P37482 bacillus su
46	29	67.4	463	1	DALD_YERPE	P58709 yersinia pe
47	29	67.4	464	1	CISY_PIG	P00889 sus scrofa
48	29	67.4	466	1	CISY_HUMAN	O75390 homo sapien
49	29	67.4	492	1	NU4M_CHOCR	P48915 chondrus cr
50	29	67.4	523	1	NU4M_PROWI	Q37617 prototheca

ALIGNMENTS

RESULT 1
CTG1_HUMAN
ID CTG1_HUMAN STANDARD; PRT; 180 AA.
AC P78358;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cancer/testis antigen 1 (Autoimmunogenic cancer/testis antigen NY-ESO-
DE 1).
GN CTAG1 OR CTAG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97203161; PubMed=9050879;
RA Chen Y.-T., Scanlan M.J., Sahin U., Tuereci O., Gure A.O., Tsang S.,
RA Williamson B., Stockert E., Pfreundschuh M., Old L.J.;
RT "A testicular antigen aberrantly expressed in human cancers detected
RT by autologous antibody screening.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:1914-1918(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=98289662; PubMed=9626360;
RA Lethe B., Lucas S., Michaux L., de Smet C., Godelaine D., Serrano A.,
RA de Plaen E., Boon T.;
RT "LAGE-1, a new gene with tumor specificity.";
RL Int. J. Cancer 76:903-908(1998).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=98430682; PubMed=9759882;
RA Wang R.-F., Johnston S.L., Zeng G., Topalian S.L.,
RA Schwartzentruber D.J., Rosenberg S.A.;
RT "A breast and melanoma-shared tumor antigen: T cell responses to
RT antigenic peptides translated from different open reading frames.";
RL J. Immunol. 161:3596-3606(1998).
CC -!- TISSUE SPECIFICITY: Expressed in testis and ovary and in a wide
CC variety of cancers. Detected in uterine myometrium.
CC -!- SIMILARITY: Belongs to the CTAG family.
CC ----------

CC ----------
DR EMBL; U87459; AAB49693.1; -.

```
DR   EMBL; AJ003149; CAA05908.1; -.
DR   EMBL; AF038567; AAD05202.1; -.
DR   Genew; HGNC:2491; CTAG1.
DR   MIM; 300156; -.
KW   Transmembrane; Antigen.
FT   DOMAIN        5     82       GLY-RICH.
FT   TRANSMEM    156    172       POTENTIAL.
SQ   SEQUENCE   180 AA;  17992 MW;  B122C5C2C8BE1569 CRC64;

  Query Match             97.7%;  Score 42;  DB 1;  Length 180;
  Best Local Similarity   100.0%;  Pred. No. 0.69;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              ||||||||
Db        157 SLLMWITQ 164


RESULT 2
YM10_YEAST
ID   YM10_YEAST     STANDARD;      PRT;   122 AA.
AC   Q03880;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical 14.0 kDa protein in RPL15B-GCR3 intergenic region.
GN   YMR123W OR YM8564.05.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313268; PubMed=9169872;
RA   Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA   Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA   Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA   Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   XIII.";
RL   Nature 387:90-93(1997).
CC   -!- SIMILARITY: TO S.POMBE SPBP23A10.02.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z49273; CAA89272.1; -.
DR   PIR; S54492; S54492.
DR   GermOnline; 142792; -.
DR   SGD; S0004730; PKR1.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     21     41       POTENTIAL.
FT   TRANSMEM     47     67       POTENTIAL.
SQ   SEQUENCE   122 AA;  13962 MW;  5CAA6929D55F2C40 CRC64;

  Query Match             79.1%;  Score 34;  DB 1;  Length 122;
  Best Local Similarity   85.7%;  Pred. No. 13;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWIT 7
              |||:|||
Db         56 SLLLWIT 62


RESULT 3
XT10_ARATH
ID   XT10_ARATH     STANDARD;      PRT;   299 AA.
AC   Q9ZVK1;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Probable xyloglucan endotransglucosylase/hydrolase protein 10
DE   precursor (EC 2.4.1.207) (At-XTH10) (XTH-10).
GN   XTH10 OR XTR14 OR AT2G14620 OR T6B13.14.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA   Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA   Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA   Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA   Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA   Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=22954850; PubMed=14593172;
RA   Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA   Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA   Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA   Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA   Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA   Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA   Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA   Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
RA   Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA   Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA   Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
RA   Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT   "Empirical analysis of transcriptional activity in the Arabidopsis
RT   genome.";
RL   Science 302:842-846(2003).
RN   [3]
RP   NOMENCLATURE.
RX   MEDLINE=22402747; PubMed=12514239;
RA   Rose J.K.C., Braam J., Fry S.C., Nishitani K.;
RT   "The XTH family of enzymes involved in xyloglucan
RT   endotransglucosylation and endohydrolysis: current perspectives and a
RT   new unifying nomenclature.";
RL   Plant Cell Physiol. 43:1421-1435(2002).
CC   -!- FUNCTION: Catalyzes xyloglucan endohydrolysis (XEH) and/or
CC       endotransglycosylation (XET). Cleaves and religates xyloglucan
CC       polymers, an essential constituant of the primary cell wall, and
CC       thereby participates in cell wall construction of growing tissues
CC       (By similarity).
CC   -!- CATALYTIC ACTIVITY: Breaks a beta-(1->4) bond in the backbone of a
CC       xyloglucan and transfers the xyloglucanyl segment on to O-4 of the
CC       non-reducing terminal glucose residue of an acceptor, which can be
CC       a xyloglucan or an oligosaccharide of xyloglucan.
CC   -!- SUBCELLULAR LOCATION: Apoplast (Probable).
CC   -!- PTM: Contains at least one intrachain disulfide bond essential for
CC       its enzymatic activity (By similarity).
CC   -!- SIMILARITY: Belongs to family 16 of glycosyl hydrolases. XTH
CC       goup 1 subfamily.
CC   -!- DATABASE: NAME=XTH-World;
CC       WWW="http://www.plantbio.cornell.edu/XTH".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AC005398; AAC69380.1; -.
DR   EMBL; AY070415; AAL49911.1; -.
DR   EMBL; AY096596; AAM20246.1; -.
DR   PIR; D84519; D84519.
DR   HSSP; P23904; 1AJK.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000757; Glyco_hydro_16.
DR   Pfam; PF00722; Glyco_hydro_16; 1.
DR   PROSITE; PS01034; GLYCOSYL_HYDROL_F16; FALSE_NEG.
KW   Hydrolase; Transferase; Glycosidase; Cell wall; Apoplast; Signal;
KW   Multigene family.
FT   SIGNAL        1     29       POTENTIAL.
FT   CHAIN        30    299       PROBABLE XYLOGLUCAN
FT                                ENDOTRANSGLUCOSYLASE/HYDROLASE PROTEIN
FT                                10.
FT   ACT_SITE    111    111       NUCLEOPHILE (BY SIMILARITY).
FT   ACT_SITE    115    115       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD     51     51       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    238    238       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   299 AA;  34687 MW;  C8A688EFB7E910A5 CRC64;

  Query Match             79.1%;  Score 34;  DB 1;  Length 299;
  Best Local Similarity   62.5%;  Pred. No. 31;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              |||:|::|
Db         21 SLLLWVSQ 28


RESULT 4
YD54_METJA
ID   YD54_METJA     STANDARD;      PRT;   145 AA.
AC   Q58749;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein MJ1354.
GN   MJ1354.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U67575; AAB99366.1; -.
DR   PIR; A64469; A64469.
DR   TIGR; MJ1354; -.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM      3     23       POTENTIAL.
FT   TRANSMEM     83    103       POTENTIAL.
FT   TRANSMEM    105    125       POTENTIAL.
SQ   SEQUENCE   145 AA;  16713 MW;  15EFEE71C5262B37 CRC64;

  Query Match             74.4%;  Score 32;  DB 1;  Length 145;
  Best Local Similarity   71.4%;  Pred. No. 35;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWIT 7
              |||:|:|
Db        106 SLLVWVT 112


RESULT 5
P85A_BOVIN
ID   P85A_BOVIN     STANDARD;      PRT;   724 AA.
AC   P23727;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phosphatidylinositol 3-kinase regulatory alpha subunit (PI3-kinase
DE   p85-alpha subunit) (PtdIns-3-kinase p85-alpha) (PI3K).
GN   PIK3R1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91191567; PubMed=1707345;
RA   Otsu M., Hiles I.D., Goot I., Fry M.J., Ruiz-Larrea F., Panayotou G.,
RA   Thompson A., Dhand R., Hsuan J., Totty N., Smith A.D., Morgan S.J.,
RA   Courtneidge S.A., Parker P.J., Waterfield M.D.;
RT   "Characterization of two 85 kd proteins that associate with receptor
RT   tyrosine kinases, middle-T/pp60c-src complexes, and PI3-kinase.";
RL   Cell 65:91-104(1991).
RN   [2]
RP   CIRCULAR DICHROISM AND FLUORESCENCE SPECTROSCOPY.
RX   MEDLINE=93049176; PubMed=1330535;
RA   Panayotou G., Bax B., Gout I., Federwisch M., Wroblowski B., Dhand R.,
RA   Fry M.J., Blundell T.L., Wollmer A., Waterfield M.D.;
RT   "Interaction of the p85 subunit of PI 3-kinase and its N-terminal SH2
RT   domain with a PDGF receptor phosphorylation site: structural features
RT   and analysis of conformational changes.";
RL   EMBO J. 11:4261-4272(1992).
RN   [3]
RP   STRUCTURE BY NMR OF 1-84.
RX   MEDLINE=93272320; PubMed=7684655;
RA   Booker G.W., Gout I., Downing A.K., Driscoll P.C., Boyd J.,
RA   Waterfield M.D., Campbell I.D.;
RT   "Solution structure and ligand-binding site of the SH3 domain of the
RT   p85 alpha subunit of phosphatidylinositol 3-kinase.";
RL   Cell 73:813-822(1993).
RN   [4]
RP   STRUCTURE BY NMR OF 314-431.
RX   MEDLINE=92357146; PubMed=1323062;
RA   Booker G.W., Breeze A.L., Downing A.K., Panayotou G., Gout I.,
RA   Waterfield M.D., Campbell I.D.;
RT   "Structure of an SH2 domain of the p85 alpha subunit of
RT   phosphatidylinositol-3-OH kinase.";
RL   Nature 358:684-687(1992).
RN   [5]
RP   STRUCTURE BY NMR OF 321-434.
RX   MEDLINE=97110350; PubMed=8952511;
RA   Guenther U.L., Liu Y., Sanford D., Bachovchin W.W., Schaffhausen B.;
RT   "NMR analysis of interactions of a phosphatidylinositol 3'-kinase SH2
```

```
RT   domain with phosphotyrosine peptides reveals interdependence of major
RT   binding sites.";
RL   Biochemistry 35:15570-15581(1996).
RN   [6]
RP   STRUCTURE BY NMR OF 614-724.
RX   MEDLINE=98173872; PubMed=9512716;
RA   Siegal G., Davis B., Kristensen S.M., Sankar A., Linacre J.,
RA   Stein R.C., Panayotou G., Waterfield M.D., Driscoll P.C.;
RT   "Solution structure of the C-terminal SH2 domain of the p85 alpha
RT   regulatory subunit of phosphoinositide 3-kinase.";
RL   J. Mol. Biol. 276:461-478(1998).
CC   -!- FUNCTION: BINDS TO ACTIVATED (PHOSPHORYLATED) PROTEIN-TYROSINE
CC       KINASES, THROUGH ITS SH2 DOMAIN, AND ACTS AS AN ADAPTER, MEDIATING
CC       THE ASSOCIATION OF THE P110 CATALYTIC UNIT TO THE PLASMA MEMBRANE.
CC       NECESSARY FOR THE INSULIN-STIMULATED INCREASE IN GLUCOSE UPTAKE
CC       AND GLYCOGEN SYNTHESIS IN INSULIN-SENSITIVE TISSUES.
CC   -!- SUBUNIT: HETERODIMER OF A P110 (CATALYTIC) AND A P85 (REGULATORY)
CC       SUBUNITS.
CC   -!- SIMILARITY: Belongs to the PI3K p85 subunit family.
CC   -!- SIMILARITY: Contains 1 Rho-GAP domain.
CC   -!- SIMILARITY: Contains 2 SH2 domains.
CC   -!- SIMILARITY: Contains 1 SH3 domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M61745; AAA79511.1; -.
DR   PIR; A38749; A38749.
DR   PDB; 2PNA; 31-JAN-94.
DR   PDB; 2PNB; 31-JAN-94.
DR   PDB; 2PNI; 31-OCT-93.
DR   PDB; 1PNJ; 31-OCT-93.
DR   PDB; 1BFI; 25-FEB-98.
DR   PDB; 1BFJ; 25-FEB-98.
DR   PDB; 1QAD; 26-SEP-01.
DR   InterPro; IPR001720; PI3kinase_P85.
DR   InterPro; IPR008936; Rho_GAP.
DR   InterPro; IPR000198; RhoGAP.
DR   InterPro; IPR000980; SH2.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00620; RhoGAP; 1.
DR   Pfam; PF00017; SH2; 2.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00678; PI3KINASEP85.
DR   PRINTS; PR00401; SH2DOMAIN.
DR   ProDom; PD000093; SH2; 2.
DR   SMART; SM00324; RhoGAP; 1.
DR   SMART; SM00252; SH2; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50238; RHOGAP; 1.
DR   PROSITE; PS50001; SH2; 2.
DR   PROSITE; PS50002; SH3; 1.
KW   SH3 domain; SH2 domain; Repeat; 3D-structure.
FT   DOMAIN        3     79       SH3.
FT   DOMAIN      113    301       RHO-GAP.
FT   DOMAIN      333    428       SH2 1.
FT   DOMAIN      624    718       SH2 2.
FT   STRAND        7     10
FT   STRAND       14     14
FT   TURN         19     20
FT   STRAND       21     21
FT   STRAND       24     24
FT   TURN         26     27
FT   STRAND       29     31
FT   HELIX        37     39
FT   TURN         40     41
FT   TURN         50     52
FT   STRAND       55     60
FT   TURN         61     64
FT   STRAND       65     70
FT   STRAND       74     78
FT   TURN         81     82
FT   HELIX       340    347
FT   TURN        348    348
FT   STRAND      354    359
FT   STRAND      368    373
FT   STRAND      378    382
FT   STRAND      384    384
FT   STRAND      391    391
FT   HELIX       401    409
FT   STRAND      413    413
FT   TURN        418    419
FT   STRAND      422    422
FT   STRAND      427    427
FT   HELIX       617    619
FT   HELIX       621    624
FT   STRAND      625    625
FT   TURN        627    628
FT   HELIX       631    638
FT   TURN        639    640
FT   TURN        643    644
FT   STRAND      645    650
FT   STRAND      657    662
FT   STRAND      667    672
FT   STRAND      675    675
FT   TURN        676    677
FT   STRAND      678    678
FT   STRAND      682    683
FT   TURN        684    685
FT   STRAND      686    687
FT   HELIX       691    700
FT   TURN        703    703
FT   STRAND      704    704
FT   TURN        705    705
FT   HELIX       706    708
FT   TURN        709    709
FT   STRAND      710    710
FT   STRAND      716    717
FT   TURN        718    719
SQ   SEQUENCE   724 AA;  83497 MW;  EBDF6E754BBF7321 CRC64;

  Query Match             74.4%;  Score 32;  DB 1;  Length 724;
  Best Local Similarity   83.3%;  Pred. No. 1.7e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 LMWITQ 8
              |||:||
Db        581 LMWLTQ 586


RESULT 6
P85A_HUMAN
ID   P85A_HUMAN     STANDARD;      PRT;   724 AA.
AC   P27986;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phosphatidylinositol 3-kinase regulatory alpha subunit (PI3-kinase
DE   p85-alpha subunit) (PtdIns-3-kinase p85-alpha) (PI3K).
GN   PIK3R1 OR GRB1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91191565; PubMed=1849461;
RA   Skolnik E.Y., Margolis B., Mohammadi M., Lowenstein E., Fischer R.,
RA   Drepps A., Ullrich A., Schlessinger J.;
RT   "Cloning of PI3 kinase-associated p85 utilizing a novel method for
RT   expression/cloning of target proteins for receptor tyrosine
```

```
RT   kinases.";
RL   Cell 65:83-90(1991).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 1-85.
RX   MEDLINE=96196433; PubMed=8648629;
RA   Liang J., Chen J.K., Schreiber S.L., Clardy J.;
RT   "Crystal structure of PI3K SH3 domain at 2.0-A resolution.";
RL   J. Mol. Biol. 257:632-643(1996).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 324-434.
RX   MEDLINE=96185448; PubMed=8599763;
RA   Nolte R.T., Eck M.J., Schlessinger J., Shoelson S.E., Harrison S.C.;
RT   "Crystal structure of the PI 3-kinase p85 amino-terminal SH2 domain
RT   and its phosphopeptide complexes.";
RL   Nat. Struct. Biol. 3:364-373(1996).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 115-298.
RX   MEDLINE=97121392; PubMed=8962058;
RA   Musacchio A., Cantley L.C., Harrison S.C.;
RT   "Crystal structure of the breakpoint cluster region-homology domain
RT   from phosphoinositide 3-kinase p85 alpha subunit.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:14373-14378(1996).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (1.79 ANGSTROMS) OF 617-724 IN COMPLEX WITH
RP   PDGFRB.
RX   MEDLINE=21450159; PubMed=11567151;
RA   Pauptit R.A., Dennis C.A., Derbyshire D.J., Breeze A.L., Weston S.A.,
RA   Rowsell S., Murshudov G.N.;
RT   "NMR trial models: experiences with the colicin immunity protein Im7
RT   and the p85alpha C-terminal SH2-peptide complex.";
RL   Acta Crystallogr. D 57:1397-1404(2001).
RN   [6]
RP   STRUCTURE BY NMR OF 1-79.
RX   MEDLINE=93208889; PubMed=7681364;
RA   Koyama S., Yu H., Dalgarno D.C., Shin T.B., Zydowsky L.D.,
RA   Schreiber S.L.;
RT   "Structure of the PI3K SH3 domain and analysis of the SH3 family.";
RL   Cell 72:945-952(1993).
RN   [7]
RP   STRUCTURE BY NMR OF 91-104.
RX   MEDLINE=97121261; PubMed=8961927;
RA   Renzoni D.A., Pugh D.J., Siligardi G., Das P., Morton C.J., Rossi C.,
RA   Waterfield M.D., Campbell I.D., Ladbury J.E.;
RT   "Structural and thermodynamic characterization of the interaction of
RT   the SH3 domain from Fyn with the proline-rich binding site on the p85
RT   subunit of PI3-kinase.";
RL   Biochemistry 35:15646-15653(1996).
RN   [8]
RP   STRUCTURE BY NMR OF 617-724.
RX   MEDLINE=96312955; PubMed=8670861;
RA   Breeze A.L., Kara B.V., Barratt D.G., Anderson M., Smith J.C.,
RA   Luke R.W., Best J.R., Cartlidge S.A.;
RT   "Structure of a specific peptide complex of the carboxy-terminal SH2
RT   domain from the p85 alpha subunit of phosphatidylinositol 3-kinase.";
RL   EMBO J. 15:3579-3589(1996).
RN   [9]
RP   VARIANT ILE-326.
RX   MEDLINE=97184306; PubMed=9032108;
RA   Hansen T., Andersen C.B., Echwald S.M., Urhammer S.A., Clausen J.O.,
RA   Vestergaard H., Owens D., Hansen L., Pedersen O.;
RT   "Identification of a common amino acid polymorphism in the p85alpha
RT   regulatory subunit of phosphatidylinositol 3-kinase: effects on
RT   glucose disappearance constant, glucose effectiveness, and the
RT   insulin sensitivity index.";
RL   Diabetes 46:494-501(1997).
RN   [10]
RP   VARIANT SEVERE INSULIN RESISTANCE GLN-409, AND VARIANT ILE-326.
RX   MEDLINE=20230645; PubMed=10768093;
RA   Baynes K.C.R., Beeton C.A., Panayotou G., Stein R., Soos M.,
RA   Hansen T., Simpson H., O'Rahilly S., Shepherd P.R., Whitehead J.P.;
RT   "Natural variants of human p85 alpha phosphoinositide 3-kinase in
RT   severe insulin resistance: a novel variant with impaired
RT   insulin-stimulated lipid kinase activity.";
RL   Diabetologia 43:321-331(2000).
CC   -!- FUNCTION: Binds to activated (phosphorylated) protein-Tyr kinases,
CC       through its SH2 domain, and acts as an adapter, mediating the
CC       association of the P110 catalytic unit to the plasma membrane.
CC       Necessary for the insulin-stimulated increase in glucose uptake
CC       and glycogen synthesis in insulin-sensitive tissues.
CC   -!- SUBUNIT: Heterodimer of a P110 (catalytic) and a P85 (regulatory)
CC       subunits. Interacts with phosphorylated TOM1L1 (By similarity).
CC   -!- DISEASE: Defects in PIK3R1 are a cause of severe insulin
CC       resistance.
CC   -!- SIMILARITY: Belongs to the PI3K p85 subunit family.
CC   -!- SIMILARITY: Contains 1 Rho-GAP domain.
CC   -!- SIMILARITY: Contains 2 SH2 domains.
CC   -!- SIMILARITY: Contains 1 SH3 domain.
CC   -!- DATABASE: NAME=PROW; NOTE=PROW 1:6-12(2000);
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/guide/1773542685_g.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M61906; -; NOT_ANNOTATED_CDS.
DR   PIR; A38748; A38748.
DR   PDB; 1PKS; 31-MAY-94.
DR   PDB; 1PKT; 31-MAY-94.
DR   PDB; 1PHT; 07-DEC-95.
DR   PDB; 1PBW; 12-MAR-97.
DR   PDB; 1PIC; 17-SEP-97.
DR   PDB; 1A0N; 25-FEB-98.
DR   PDB; 1AZG; 25-FEB-98.
DR   PDB; 1H9O; 27-NOV-01.
DR   Genew; HGNC:8979; PIK3R1.
DR   MIM; 171833; -.
DR   GO; GO:0016303; F:phosphatidylinositol 3-kinase activity; NAS.
DR   GO; GO:0005545; F:phosphatidylinositol binding; NAS.
DR   GO; GO:0007242; P:intracellular signaling cascade; NAS.
DR   InterPro; IPR001720; PI3kinase_P85.
DR   InterPro; IPR008936; Rho_GAP.
DR   InterPro; IPR000198; RhoGAP.
DR   InterPro; IPR000980; SH2.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00620; RhoGAP; 1.
DR   Pfam; PF00017; SH2; 2.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00678; PI3KINASEP85.
DR   PRINTS; PR00401; SH2DOMAIN.
DR   ProDom; PD000093; SH2; 2.
DR   SMART; SM00324; RhoGAP; 1.
DR   SMART; SM00252; SH2; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50238; RHOGAP; 1.
DR   PROSITE; PS50001; SH2; 2.
DR   PROSITE; PS50002; SH3; 1.
KW   SH3 domain; SH2 domain; Repeat; 3D-structure; Polymorphism;
KW   Disease mutation.
FT   DOMAIN        3     79       SH3.
FT   DOMAIN      113    301       RHO-GAP.
FT   DOMAIN      333    428       SH2 1.
FT   DOMAIN      624    718       SH2 2.
FT   VARIANT     326    326       M -> I.
FT                                /FTId=VAR_010023.
FT   VARIANT     409    409       R -> Q (in severe insulin resistance;
FT                                reduction of insulin-stimulated
FT                                activity).
FT                                /FTId=VAR_010024.
FT   STRAND        4     10
FT   STRAND       14     14
FT   TURN         19     20
FT   STRAND       21     21
```

```
FT   STRAND       24     24
FT   TURN         26     27
FT   STRAND       29     33
FT   HELIX        34     38
FT   TURN         39     41
FT   TURN         44     45
FT   HELIX        46     48
FT   HELIX        50     53
FT   STRAND       55     60
FT   TURN         61     64
FT   STRAND       65     70
FT   HELIX        71     73
FT   STRAND       74     81
FT   HELIX       118    121
FT   TURN        124    125
FT   HELIX       130    143
FT   TURN        144    145
FT   TURN        147    150
FT   HELIX       158    165
FT   TURN        167    168
FT   HELIX       174    176
FT   STRAND      177    177
FT   HELIX       179    192
FT   HELIX       200    208
FT   TURN        209    212
FT   HELIX       216    227
FT   TURN        228    228
FT   TURN        230    231
FT   TURN        234    234
FT   HELIX       235    251
FT   TURN        252    253
FT   HELIX       254    257
FT   HELIX       261    272
FT   TURN        273    274
FT   TURN        277    278

  Query Match             74.4%;  Score 32;  DB 1;  Length 724;
  Best Local Similarity   83.3%;  Pred. No. 1.7e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 LMWITQ 8
              |||:||
Db        581 LMWLTQ 586


RESULT 7
P85A_MOUSE
ID   P85A_MOUSE     STANDARD;      PRT;   724 AA.
AC   P26450;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Phosphatidylinositol 3-kinase regulatory alpha subunit (PI3-kinase
DE   p85-alpha subunit) (PtdIns-3-kinase p85-alpha) (PI3K).
GN   PIK3R1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=BALB/c;
RX   MEDLINE=91191564; PubMed=1849460;
RA   Escobedo J.A., Navankasattusas S., Kavanaugh W.M., Milfay D.,
RA   Fried V.A., Williams L.T.;
RT   "cDNA cloning of a novel 85 kd protein that has SH2 domains and
RT   regulates binding of PI3-kinase to the PDGF beta-receptor.";
RL   Cell 65:75-82(1991).
RN   [2]
RP   INTERACTION WITH TOM1L1.
RX   MEDLINE=21659738; PubMed=11711534;
RA   Seykora J.T., Mei L., Dotto G.P., Stein P.L.;
RT   "'Srcasm: a novel Src activating and signaling molecule.";
RL   J. Biol. Chem. 277:2812-2822(2002).
CC   -!- FUNCTION: Binds to activated (phosphorylated) protein-Tyr kinases,
CC       through its SH2 domain, and acts as an adapter, mediating the
CC       association of the P110 catalytic unit to the plasma membrane.
CC       Necessary for the insulin-stimulated increase in glucose uptake
CC       and glycogen synthesis in insulin-sensitive tissues.
CC   -!- SUBUNIT: Heterodimer of a P110 (catalytic) and a P85 (regulatory)
CC       subunits. Interacts with phosphorylated TOM1L1.
CC   -!- SIMILARITY: Belongs to the PI3K p85 subunit family.
CC   -!- SIMILARITY: Contains 1 Rho-GAP domain.
CC   -!- SIMILARITY: Contains 2 SH2 domains.
CC   -!- SIMILARITY: Contains 1 SH3 domain.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M60651; AAA39886.1; -.
DR   HSSP; P23727; 1BFI.
DR   MGD; MGI:97583; Pik3r1.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0030183; P:B-cell differentiation; IMP.
DR   InterPro; IPR001720; PI3kinase_P85.
DR   InterPro; IPR008936; Rho_GAP.
DR   InterPro; IPR000198; RhoGAP.
DR   InterPro; IPR000980; SH2.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00620; RhoGAP; 1.
DR   Pfam; PF00017; SH2; 2.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00678; PI3KINASEP85.
DR   PRINTS; PR00401; SH2DOMAIN.
DR   ProDom; PD000093; SH2; 2.
DR   SMART; SM00324; RhoGAP; 1.
DR   SMART; SM00252; SH2; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50238; RHOGAP; 1.
DR   PROSITE; PS50001; SH2; 2.
DR   PROSITE; PS50002; SH3; 1.
KW   SH3 domain; SH2 domain; Repeat.
FT   DOMAIN        3     79       SH3.
FT   DOMAIN      113    301       RHO-GAP.
FT   DOMAIN      333    428       SH2 1.
FT   DOMAIN      624    718       SH2 2.
SQ   SEQUENCE   724 AA;  83414 MW;  DEAAE6EE297CF07A CRC64;

  Query Match             74.4%;  Score 32;  DB 1;  Length 724;
  Best Local Similarity   83.3%;  Pred. No. 1.7e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 LMWITQ 8
              |||:||
Db        581 LMWLTQ 586


RESULT 8
P85A_RAT
ID   P85A_RAT       STANDARD;      PRT;   724 AA.
AC   Q63787; O55085; P70544; Q63790;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phosphatidylinositol 3-kinase regulatory alpha subunit (PI3-kinase
DE   p85-alpha subunit) (PtdIns-3-kinase p85-alpha) (PI3K).
GN   PIK3R1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
```

```
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS P85-ALPHA AND P55-ALPHA).
RC   STRAIN=Wistar; TISSUE=Brain;
RX   MEDLINE=96214979; PubMed=8621382;
RA   Inukai K., Anai M., van Breda E., Hosaka T., Katagiri H., Funaki M.,
RA   Fukushima Y., Ogihara T., Yazaki Y., Kikuchi M., Oka Y., Asano T.;
RT   "A novel 55-kDa regulatory subunit for phosphatidylinositol 3-kinase
RT   structurally similar to p55PIK Is generated by alternative splicing
RT   of the p85alpha gene.";
RL   J. Biol. Chem. 271:5317-5320(1996).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM P50-ALPHA).
RC   TISSUE=Liver;
RX   MEDLINE=97218222; PubMed=9065454;
RA   Inukai K., Funaki M., Ogihara T., Katagiri H., Kanda A., Anai M.,
RA   Fukushima Y., Hosaka T., Suzuki M., Shin B., Takata K., Yazaki Y.,
RA   Kikuchi M., Oka Y., Asano T.;
RT   "p85alpha gene generates three isoforms of regulatory subunit for
RT   phosphatidylinositol 3-kinase (PI 3-Kinase), p50alpha, p55alpha, and
RT   p85alpha, with different PI 3-kinase activity elevating responses to
RT   insulin.";
RL   J. Biol. Chem. 272:7873-7882(1997).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM P50-ALPHA).
RC   TISSUE=Liver;
RX   MEDLINE=97079666; PubMed=8921377;
RA   Fruman D.A., Cantley L.C., Carpenter C.L.;
RT   "Structural organization and alternative splicing of the murine
RT   phosphoinositide 3-kinase p85 alpha gene.";
RL   Genomics 37:113-121(1996).
CC   -!- FUNCTION: Binds to activated (phosphorylated) protein-Tyr kinases,
CC       through its SH2 domain, and acts as an adapter, mediating the
CC       association of the P110 catalytic unit to the plasma membrane.
CC       Necessary for the insulin-stimulated increase in glucose uptake
CC       and glycogen synthesis in insulin-sensitive tissues.
CC   -!- SUBUNIT: Heterodimer of a P110 (catalytic) and a P85 (regulatory)
CC       subunits. Interacts with phosphorylated TOM1L1 (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=p85-alpha;
CC         IsoId=Q63787-1; Sequence=Displayed;
CC       Name=p55-alpha;
CC         IsoId=Q63787-2; Sequence=VSP_004709, VSP_004710;
CC       Name=p50-alpha;
CC         IsoId=Q63787-3; Sequence=VSP_004711, VSP_004712;
CC   -!- TISSUE SPECIFICITY: The P85-alpha isoform is widely expressed.
CC       Expression of the P55-alpha isoform is highest in brain and
CC       skeletal muscle. The P50-alpha isoform is abundant in liver with
CC       lower levels in brain and muscle.
CC   -!- SIMILARITY: Belongs to the PI3K p85 subunit family.
CC   -!- SIMILARITY: Contains 1 Rho-GAP domain.
CC   -!- SIMILARITY: Contains 2 SH2 domains.
CC   -!- SIMILARITY: Contains 1 SH3 domain.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D64045; BAA18932.1; -.
DR   EMBL; D64048; BAA18933.1; -.
DR   EMBL; U50412; AAC52846.1; -.
DR   EMBL; D78486; BAA24426.1; -.
DR   PDB; 1FU5; 21-FEB-01.
DR   PDB; 1FU6; 21-FEB-01.
DR   InterPro; IPR001720; PI3kinase_P85.
DR   InterPro; IPR008936; Rho_GAP.
DR   InterPro; IPR000198; RhoGAP.
DR   InterPro; IPR000980; SH2.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00620; RhoGAP; 1.
DR   Pfam; PF00017; SH2; 2.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00678; PI3KINASEP85.
DR   PRINTS; PR00401; SH2DOMAIN.
DR   ProDom; PD000093; SH2; 2.
DR   SMART; SM00324; RhoGAP; 1.
DR   SMART; SM00252; SH2; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50238; RHOGAP; 1.
DR   PROSITE; PS50001; SH2; 2.
DR   PROSITE; PS50002; SH3; 1.
KW   SH3 domain; SH2 domain; Repeat; Alternative splicing; 3D-structure.
FT   DOMAIN        3     79       SH3.
FT   DOMAIN      113    301       RHO-GAP.
FT   DOMAIN      333    428       SH2 1.
FT   DOMAIN      624    718       SH2 2.
FT   VARSPLIC      1    270       Missing (in isoform p55-alpha).
FT                                /FTId=VSP_004709.
FT   VARSPLIC    271    304       VLFRFPAASSDNTEHLIKAVELLISAEWSERQPA -> MYT
FT                                TVWTMEDLDLECAKTDINCGTDLMFYIEMDP (in
FT                                isoform p55-alpha).
FT                                /FTId=VSP_004710.
FT   VARSPLIC      1    300       Missing (in isoform p50-alpha).
FT                                /FTId=VSP_004711.
FT   VARSPLIC    301    306       RQPAPA -> MHNLQT (in isoform p50-alpha).
FT                                /FTId=VSP_004712.
SQ   SEQUENCE   724 AA;  83531 MW;  95C65CF612873B84 CRC64;

  Query Match             74.4%;  Score 32;  DB 1;  Length 724;
  Best Local Similarity   83.3%;  Pred. No. 1.7e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 LMWITQ 8
              |||:||
Db        581 LMWLTQ 586


RESULT 9
RRPL_MABVM
ID   RRPL_MABVM     STANDARD;      PRT;  2331 AA.
AC   P31352;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   RNA-directed RNA polymerase (EC 2.7.7.48) (Large structural protein)
DE   (L protein).
GN   L.
OS   Marburg virus (strain Musoke).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales; Filoviridae;
OC   Marburg-like viruses.
OX   NCBI_TaxID=33727;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92188528; PubMed=1546452;
RA   Muehlberger E., Sanchez A., Randolf A., Will C., Kiley M.P.,
RA   Klenk H.D., Feldmann H.;
RT   "The nucleotide sequence of the L gene of Marburg virus, a filovirus:
RT   homologies with paramyxoviruses and rhabdoviruses.";
RL   Virology 187:534-547(1992).
RN   [2]
RP   REVISIONS.
RA   Feldmann H.;
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN IS PROBABLY A COMPONENT OF THE ACTIVE
CC       POLYMERASE. IT MAY FUNCTION IN RNA SYNTHESIS, CAPPING, AS WELL AS
CC       METHYLATION OF CAPS, AND POLY(A) SYNTHESIS.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SIMILARITY: PRIMARILY WITH THE N-TERMINAL HALF OF THE L PROTEINS
CC       OF RHABDOVIRUSES AND PARAMYXOVIRUSES.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M92834; AAA46562.1; ALT_SEQ.
DR   EMBL; Z12132; CAA78120.1; -.
DR   PIR; A42450; RRIWMV.
DR   InterPro; IPR007098; RNA_pol_monon.
DR   InterPro; IPR001016; Viral_RNA_pol_L.
DR   Pfam; PF00946; Paramyx_RNA_pol; 1.
KW   Transferase; RNA-directed RNA polymerase.
SQ   SEQUENCE   2331 AA;  267090 MW;  1D54C60DA1BED3CF CRC64;

  Query Match             74.4%;  Score 32;  DB 1;  Length 2331;
  Best Local Similarity   62.5%;  Pred. No. 5.4e+02;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              | |:|:||
Db       1249 SRLLWVTQ 1256


RESULT 10
RRPL_MABVP
ID   RRPL_MABVP     STANDARD;      PRT;  2331 AA.
AC   P35262;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   RNA-directed RNA polymerase (EC 2.7.7.48) (Large structural protein)
DE   (L protein).
GN   L.
OS   Marburg virus (strain Popp).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales; Filoviridae;
OC   Marburg-like viruses.
OX   NCBI_TaxID=33728;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96028047; PubMed=7487490;
RA   Bukreyev A.A., Volchkov V.E., Blinov V.M., Dryga S.A., Netesov S.V.;
RT   "The complete nucleotide sequence of the Popp (1967) strain of Marburg
RT   virus: a comparison with the Musoke (1980) strain.";
RL   Arch. Virol. 140:1589-1600(1995).
CC   -!- FUNCTION: THIS PROTEIN IS PROBABLY A COMPONENT OF THE ACTIVE
CC       POLYMERASE. IT MAY FUNCTION IN RNA SYNTHESIS, CAPPING, AS WELL AS
CC       METHYLATION OF CAPS, AND POLY(A) SYNTHESIS.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SIMILARITY: PRIMARILY WITH THE N-TERMINAL HALF OF THE L PROTEINS
CC       OF RHABDOVIRUSES AND PARAMYXOVIRUSES.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X68494; CAA48508.1; -.
DR   EMBL; Z29337; CAA82542.1; -.
DR   PIR; S44054; S44054.
DR   InterPro; IPR007098; RNA_pol_monon.
DR   InterPro; IPR001016; Viral_RNA_pol_L.
DR   Pfam; PF00946; Paramyx_RNA_pol; 1.
KW   Transferase; RNA-directed RNA polymerase.
SQ   SEQUENCE   2331 AA;  266635 MW;  7ED1D61D0EEF9B8B CRC64;

  Query Match             74.4%;  Score 32;  DB 1;  Length 2331;
  Best Local Similarity   62.5%;  Pred. No. 5.4e+02;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              | |:|:||
Db       1249 SRLLWVTQ 1256


RESULT 11
NUKM_SOLTU
ID   NUKM_SOLTU     STANDARD;      PRT;  213 AA.
AC   Q43844;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   NADH-ubiquinone oxidoreductase 20 kDa subunit, mitochondrial precursor
DE   (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-20KD) (CI-20KD).
OS   Solanum tuberosum (Potato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   lamiids; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4113;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Desiree; TISSUE=Leaf;
RX   MEDLINE=97071689; PubMed=8914535;
RA   Heiser V., Grohmann L., Brennicke A.;
RT   "The plant mitochondrial 22 kDa (PSST) subunit of respiratory chain
RT   complex I is encoded by a nuclear gene with enhanced transcript
RT   levels in flowers.";
RL   Plant Mol. Biol. 31:1195-1204(1996).
CC   -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC   -!- CATALYTIC ACTIVITY: NADH + acceptor = NAD(+) + reduced acceptor.
CC   -!- COFACTOR: Binds 1 4Fe-4S cluster (Potential).
CC   -!- SUBUNIT: Complex I is composed of about 40 different subunits.
CC   -!- SIMILARITY: Belongs to the complex I 20 kDa subunit family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X96671; CAA65451.1; -.
DR   PIR; T07603; T07603.
DR   InterPro; IPR006138; Cmplx1_20kDa.
DR   InterPro; IPR006137; Oxidored_q6.
DR   Pfam; PF01058; oxidored_q6; 1.
DR   PROSITE; PS01150; COMPLEX1_20K; 1.
KW   Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transit peptide;
KW   Iron-sulfur; 4Fe-4S.
FT   TRANSIT       1     31       MITOCHONDRION (POTENTIAL).
FT   CHAIN        32    213       NADH-UBIQUINONE OXIDOREDUCTASE 20 kDa
FT                                 SUBUNIT.
FT   METAL         88     88       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL         89     89       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL        153    153       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL        183    183       IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ   SEQUENCE   213 AA;  23396 MW;  8047FAB323418FCB CRC64;

  Query Match             72.1%;  Score 31;  DB 1;  Length 213;
  Best Local Similarity   85.7%;  Pred. No. 77;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LLMWITQ 8
              ||||.||
Db        207 LLMWWTQ 213


RESULT 12
CKR8_MACMU
ID   CKR8_MACMU     STANDARD;      PRT;  356 AA.
```

```
AC   O97665;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   C-C chemokine receptor type 8 (C-C CKR-8) (CC-CKR-8) (CCR-8).
GN   CCR8.
OS   Macaca mulatta (Rhesus macaque).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9544;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Spleen;
RX   MEDLINE=21354176; PubMed=11461684;
RA   Margulies B.J., Hauer D.A., Clements J.E.;
RT   "Identification and comparison of eleven rhesus macaque chemokine
RT   receptors.";
RL   AIDS Res. Hum. Retroviruses 17:981-986(2001).
CC   -!- FUNCTION: Receptor for the chemokines SCYA1/I-309, SCYA4/MIP-1-
CC       beta and SCYA17/TARC. May regulate monocyte chemotaxis and thymic
CC       cell line apoptosis (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF100205; AAC72403.1; -.
DR   InterPro; IPR004068; CC_8_receptor.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR01530; CHEMOKINER8.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein.
FT   DOMAIN        1     35       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     36     63       1 (POTENTIAL).
FT   DOMAIN       64     73       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     74     93       2 (POTENTIAL).
FT   DOMAIN       94    107       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    108    129       3 (POTENTIAL).
FT   DOMAIN      130    146       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    147    172       4 (POTENTIAL).
FT   DOMAIN      173    203       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    204    223       5 (POTENTIAL).
FT   DOMAIN      224    239       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    240    264       6 (POTENTIAL).
FT   DOMAIN      265    281       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    282    305       7 (POTENTIAL).
FT   DOMAIN      306    356       CYTOPLASMIC (POTENTIAL).
FT   DISULFID    106    184       BY SIMILARITY.
SQ   SEQUENCE   356 AA;  41210 MW;  1979628DEE44845B CRC64;

  Query Match             72.1%;  Score 31;  DB 1;  Length 356;
  Best Local Similarity   71.4%;  Pred. No. 1.3e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWIT 7
              |||:|:|
Db        155 SLLVWLT 161


RESULT 13
PFTB_BOVIN
ID   PFTB_BOVIN     STANDARD;      PRT;   437 AA.
AC   P49355; Q9TS25;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Protein farnesyltransferase beta subunit (EC 2.5.1.58) (CAAX
DE   farnesyltransferase beta subunit) (RAS proteins prenyltransferase
DE   beta) (FTase-beta).
GN   FNTB.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=93264431; PubMed=8494894;
RA   Omer C.A., Kral A.M., Diehl R.E., Prendergast G.C., Powers S.,
RA   Allen C.M., Gibbs J.B., Kohl N.E.;
RT   "Characterization of recombinant human farnesyl-protein transferase:
RT   cloning, expression, farnesyl diphosphate binding, and functional
RT   homology with yeast prenyl-protein transferases.";
RL   Biochemistry 32:5167-5176(1993).
CC   -!- FUNCTION: Catalyzes the transfer of a farnesyl moiety from
CC       farnesyl pyrophosphate to a cysteine at the fourth position from
CC       the C-terminus of several proteins. The beta subunit is
CC       responsible for peptide-binding.
CC   -!- CATALYTIC ACTIVITY: Farnesyl diphosphate + protein-cysteine = S-
CC       farnesyl protein + diphosphate.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Heterodimer of an alpha and a beta subunit.
CC   -!- SIMILARITY: Belongs to the protein prenyltransferase beta subunit
CC       family.
CC   -!- SIMILARITY: Contains 5 PFTB repeats.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L00633; AAA30524.1; -.
DR   HSSP; Q02293; 1FT2.
DR   InterPro; IPR001330; Prenyltrans.
DR   InterPro; IPR008930; Terp_cyc_toroid.
DR   Pfam; PF00432; prenyltrans; 5.
KW   Transferase; Prenyltransferase; Repeat; Zinc.
FT   REPEAT      123    164       PFTB 1.
FT   REPEAT      174    215       PFTB 2.
FT   REPEAT      222    263       PFTB 3.
FT   REPEAT      270    312       PFTB 4.
FT   REPEAT      332    374       PFTB 5.
FT   METAL       297    297       ZINC (BY SIMILARITY).
FT   METAL       299    299       ZINC (BY SIMILARITY).
FT   METAL       362    362       ZINC (BY SIMILARITY).
SQ   SEQUENCE   437 AA;  48767 MW;  CE09DFA86AC6AB64 CRC64;

  Query Match             72.1%;  Score 31;  DB 1;  Length 437;
  Best Local Similarity   71.4%;  Pred. No. 1.6e+02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SLLMWIT 7
              ||| |:|
Db        272 SLLQWVT 278


RESULT 14
PFTB_HUMAN
ID   PFTB_HUMAN     STANDARD;      PRT;   437 AA.
AC   P49356;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
```

```
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Protein farnesyltransferase beta subunit (EC 2.5.1.58) (CAAX
DE   farnesyltransferase beta subunit) (RAS proteins prenyltransferase
DE   beta) (FTase-beta).
GN   FNTB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=93264431; PubMed=8494894;
RA   Omer C.A., Kral A.M., Diehl R.E., Prendergast G.C., Powers S.,
RA   Allen C.M., Gibbs J.B., Kohl N.E.;
RT   "Characterization of recombinant human farnesyl-protein transferase:
RT   cloning, expression, farnesyl diphosphate binding, and functional
RT   homology with yeast prenyl-protein transferases.";
RL   Biochemistry 32:5167-5176(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 51-437 FROM N.A.
RC   TISSUE=Retina;
RX   MEDLINE=94102736; PubMed=8276393;
RA   Andres D.A., Milatovich A., Ozcelik T., Wenzlau J.M., Brown M.S.,
RA   Goldstein J.L., Francke U.;
RT   "cDNA cloning of the two subunits of human CAAX farnesyltransferase
RT   and chromosomal mapping of FNTA and FNTB loci and related
RT   sequences.";
RL   Genomics 18:105-112(1993).
CC   -!- FUNCTION: Catalyzes the transfer of a farnesyl moiety from
CC       farnesyl pyrophosphate to a cysteine at the fourth position from
CC       the C-terminus of several proteins. The beta subunit is
CC       responsible for peptide-binding.
CC   -!- CATALYTIC ACTIVITY: Farnesyl diphosphate + protein-cysteine = S-
CC       farnesyl protein + diphosphate.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC   -!- SUBUNIT: Heterodimer of an alpha and a beta subunit.
CC   -!- SIMILARITY: Belongs to the protein prenyltransferase beta subunit
CC       family.
CC   -!- SIMILARITY: Contains 5 PFTB repeats.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L00635; AAA35854.1; -.
DR   EMBL; BC020232; AAH20232.1; -.
DR   EMBL; L10414; AAA86286.1; -.
DR   PIR; B49274; B49274.
DR   HSSP; Q02293; 1FT2.
DR   Genew; HGNC:3785; FNTB.
DR   MIM; 134636; -.
DR   GO; GO:0004660; F:protein farnesyltransferase activity; TAS.
DR   GO; GO:0006503; P:C-terminal protein farnesylation; TAS.
DR   InterPro; IPR001330; Prenyltrans.
DR   InterPro; IPR008930; Terp_cyc_toroid.
DR   Pfam; PF00432; prenyltrans; 5.
KW   Transferase; Prenyltransferase; Repeat; Zinc.
FT   REPEAT      123    164       PFTB 1.
FT   REPEAT      174    215       PFTB 2.
FT   REPEAT      222    263       PFTB 3.
FT   REPEAT      270    312       PFTB 4.
FT   REPEAT      332    374       PFTB 5.
FT   METAL       297    297       ZINC (BY SIMILARITY).
FT   METAL       299    299       ZINC (BY SIMILARITY).
FT   METAL       362    362       ZINC (BY SIMILARITY).
FT   CONFLICT    283    283       R -> L (IN REF. 3).
SQ   SEQUENCE   437 AA;  48773 MW;  8E8E571846146709 CRC64;

  Query Match             72.1%;  Score 31;  DB 1;  Length 437;
  Best Local Similarity   71.4%;  Pred. No. 1.6e+02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SLLMWIT 7
              ||| |:|
Db        272 SLLQWVT 278


RESULT 15
PFTB_RAT
ID   PFTB_RAT       STANDARD;      PRT;   437 AA.
AC   Q02293;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Protein farnesyltransferase beta subunit (EC 2.5.1.58) (CAAX
DE   farnesyltransferase beta subunit) (RAS proteins prenyltransferase
DE   beta) (FTase-beta).
GN   FNTB.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=91309145; PubMed=1855253;
RA   Chen W.-J., Andres D.A., Goldstein J.L., Russell D.W., Brown M.S.;
RT   "cDNA cloning and expression of the peptide-binding beta subunit of
RT   rat p21ras farnesyltransferase, the counterpart of yeast DPR1/RAM1.";
RL   Cell 66:327-334(1991).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (2.25 ANGSTROMS).
RX   MEDLINE=97218306; PubMed=9065406;
RA   Park H.-W., Boduluri S.R., Mooomaw J.F., Casey P.J., Beese L.S.;
RT   "Crystal structure of protein farnesyltransferase at 2.25-A
RT   resolution.";
RL   Science 275:1800-1804(1997).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (3.4 ANGSTROMS).
RX   MEDLINE=98322062; PubMed=9657673;
RA   Long S.B., Casey P.J., Beese L.S.;
RT   "Cocrystal structure of protein farnesyltransferase complexed with a
RT   farnesyl diphosphate substrate.";
RL   Biochemistry 37:9612-9618(1998).
CC   -!- FUNCTION: Catalyzes the transfer of a farnesyl moiety from
CC       farnesyl pyrophosphate to a cysteine at the fourth position from
CC       the C-terminus of several proteins. The beta subunit is
```

```
CC       responsible for peptide-binding.
CC   -!- CATALYTIC ACTIVITY: Farnesyl diphosphate + protein-cysteine = S-
CC       farnesyl protein + diphosphate.
CC   -!- COFACTOR: Binds 1 zinc ion per subunit.
CC   -!- SUBUNIT: Heterodimer of an alpha and a beta subunit.
CC   -!- SIMILARITY: Belongs to the protein prenyltransferase beta subunit
CC       family.
CC   -!- SIMILARITY: Contains 5 PFTB repeats.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M69056; AAA41176.1; -.
DR   PIR; A40037; A40037.
DR   PDB; 1FT1; 18-MAR-98.
DR   PDB; 1FT2; 18-NOV-98.
DR   PDB; 1D8D; 09-FEB-00.
DR   PDB; 1D8E; 14-JUN-00.
DR   PDB; 1FPP; 08-JUN-99.
DR   PDB; 1FTI; 24-JAN-01.
DR   PDB; 1JCR; 02-NOV-01.
DR   PDB; 1JCS; 02-NOV-01.
DR   PDB; 1KZO; 16-OCT-02.
DR   PDB; 1KZP; 16-OCT-02.
DR   PDB; 1N94; 07-JAN-03.
DR   PDB; 1N95; 07-JAN-03.
DR   PDB; 1N9A; 07-JAN-03.
DR   PDB; 1NL4; 22-APR-03.
DR   PDB; 1QBQ; 27-MAR-00.
DR   PDB; 1QE2; 18-APR-01.
DR   PDB; 2FTI; 24-JAN-01.
DR   PDB; 3FTI; 24-JAN-01.
DR   InterPro; IPR001330; Prenyltrans.
DR   InterPro; IPR008930; Terp_cyc_toroid.
DR   Pfam; PF00432; prenyltrans; 5.
KW   Transferase; Prenyltransferase; Repeat; Zinc; 3D-structure.
FT   REPEAT      123    164       PFTB 1.
FT   REPEAT      174    215       PFTB 2.
FT   REPEAT      222    263       PFTB 3.
FT   REPEAT      270    312       PFTB 4.
FT   REPEAT      332    374       PFTB 5.
FT   METAL       297    297       ZINC.
FT   METAL       299    299       ZINC.
FT   METAL       362    362       ZINC.
FT   TURN         24     26
FT   HELIX        28     33
FT   TURN         38     39
FT   HELIX        43     64
FT   HELIX        75     85
FT   TURN         86     87
FT   STRAND       88     89
FT   HELIX        91     96
FT   TURN         97     98
FT   HELIX       100    113
FT   TURN        114    115
FT   HELIX       120    133
FT   STRAND      134    134
FT   TURN        136    137
FT   STRAND      140    141
FT   TURN        144    145
FT   STRAND      148    149
FT   HELIX       150    163
FT   HELIX       166    169
FT   TURN        170    171
FT   HELIX       174    184
FT   STRAND      185    185
FT   TURN        187    188
FT   STRAND      191    192
FT   TURN        195    196
FT   STRAND      199    199
FT   HELIX       201    214
FT   TURN        215    215
FT   TURN        219    224
FT   HELIX       225    230
FT   TURN        231    232
FT   STRAND      233    233
FT   TURN        235    236
FT   STRAND      239    239
FT   TURN        243    244
FT   HELIX       249    261
FT   TURN        262    263
FT   HELIX       265    267
FT   HELIX       270    278
FT   TURN        279    280
FT   STRAND      281    281
FT   TURN        283    285
FT   STRAND      288    289
FT   TURN        292    293
FT   STRAND      296    296
FT   TURN        298    299
FT   HELIX       300    303
FT   TURN        304    306
FT   HELIX       307    317
FT   TURN        318    319
FT   TURN        321    322
FT   HELIX       332    342
FT   STRAND      344    344
FT   TURN        346    347
FT   STRAND      350    350
FT   STRAND      353    353
FT   TURN        354    355
FT   STRAND      356    356
FT   HELIX       360    374
FT   STRAND      375    378
FT   TURN        379    380
FT   STRAND      381    384
FT   HELIX       390    392
FT   TURN        399    401
FT   HELIX       405    415
FT   TURN        416    417
FT   HELIX       424    426
FT   TURN        427    427
FT   HELIX       434    437
SQ   SEQUENCE   437 AA;  48673 MW;  41A9D6D79CD319A8 CRC64;

  Query Match             72.1%;  Score 31;  DB 1;  Length 437;
  Best Local Similarity   71.4%;  Pred. No. 1.6e+02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SLLMWIT 7
              ||| |:|
Db        272 SLLQWVT 278


RESULT 16
SNE2_HUMAN
ID   SNE2_HUMAN     STANDARD;      PRT;  6885 AA.
AC   Q8WXH0; Q8N1S3; Q8NF49; Q8TER7; Q8WWW3; Q8WWW4; Q8WWW5; Q8WXH1;
AC   Q9NU50; Q9UFQ4; Q9Y2L4; Q9Y4R1;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Nesprin 2 (Nuclear envelope spectrin repeat protein 2) (Syne-2)
DE   (Synaptic nuclear envelope protein 2) (Nucleus and actin connecting
DE   element protein) (NUANCE protein).
GN   SYNE2 OR NUA OR KIAA1011.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
```

RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 9), FUNCTION, CHARACTERIZATION, AND
RP INTERACTION WITH F-ACTIN.
RX MEDLINE=22113122; PubMed=12118075;
RA Zhen Y.-Y., Libotte T., Munck M., Noegel A.A., Korenbaum E.;
RT "NUANCE, a giant protein connecting the nucleus and actin
RT cytoskeleton.";
RL J. Cell Sci. 115:3207-3222(2002).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 4; 5 AND 7).
RX MEDLINE=21652858; PubMed=11792814;
RA Zhang Q., Skepper J.N., Yang F., Davies J.D., Hegyi L., Roberts R.G.,
RA Weissberg P.L., Ellis J.A., Shanahan C.M.;
RT "Nesprins: a novel family of spectrin-repeat-containing proteins that
RT localize to the nuclear membrane in multiple tissues.";
RL J. Cell Sci. 114:4485-4498(2001).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=22296983; PubMed=12408964;
RA Zhang Q., Ragnauth C., Greener M.J., Shanahan C.M., Roberts R.G.;
RT "The nesprins are giant actin-binding proteins, orthologous to
RT Drosophila melanogaster muscle protein MSP-300.";
RL Genomics 80:473-481(2002).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 6).
RC TISSUE=Testis;
RA Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RX PubMed=12508121;
RA Heilig R., Eckenberg R., Petit J.-L., Fonknechten N., Da Silva C.,
RA Cattolico L., Levy M., Barbe V., De Berardinis V., Ureta-Vidal A.,
RA Pelletier E., Vico V., Anthouard V., Rowen L., Madan A., Qin S.,
RA Sun H., Du H., Pepin K., Artiguenave F., Robert C., Cruaud C.,
RA Bruels T., Jaillon O., Friedlander L., Samson G., Brottier P.,
RA Cure S., Segurens B., Aniere F., Samain S., Crespeau H., Abbasi N.,
RA Aiach N., Boscus D., Dickhoff R., Dors M., Dubois I., Friedman C.,
RA Gouyvenoux M., James R., Madan A., Mairey-Estrada B., Mangenot S.,
RA Martins N., Menard M., Oztas S., Ratcliffe A., Shaffer T., Trask B.,
RA Vacherie B., Bellemere C., Belser C., Besnard-Gonnet M.,
RA Bartol-Mavel D., Boutard M., Briez-Silla S., Combette S.,
RA Dufosse-Laurent V., Ferron C., Lechaplais C., Louesse C., Muselet D.,
RA Magdelenat G., Pateau E., Petit E., Sirvain-Trukniewicz P., Trybou A.,
RA Vega-Czarny N., Bataille E., Bluet E., Bordelais I., Dubois M.,
RA Dumont C., Guerin T., Haffray S., Hammadi R., Muanga J., Pellouin V.,
RA Robert D., Wunderle E., Gauguet G., Roy A., Sainte-Marthe L.,
RA Verdier J., Verdier-Discala C., Hillier L.W., Fulton L., McPherson J.,
RA Matsuda F., Wilson R., Scarpelli C., Gyapay G., Wincker P., Saurin W.,
RA Quetier F., Waterston R., Hood L., Weissenbach J.;
RT "The DNA sequence and analysis of human chromosome 14.";
RL Nature 421:601-607(2003).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM 8).
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [7]
RP SEQUENCE FROM N.A. (ISOFORM 3), AND SEQUENCE OF 1-956 AND 5133-6885
RP FROM N.A.
RC TISSUE=Spleen, and Tongue;
RA Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O.,
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
RN [8]
RP SEQUENCE OF 5709-6885 FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosawa M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
RN [9]
RP REVISIONS.
RC TISSUE=Brain;
RX MEDLINE=22158633; PubMed=12168954;
RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT "Construction of expression-ready cDNA clones for KIAA genes: manual
RT curation of 330 KIAA cDNA clones.";
RL DNA Res. 9:99-106(2002).
RN [10]
RP SEQUENCE OF 5754-6885 FROM N.A.
RX MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansorge W., Boecher M., Bloecker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA Mewes H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Poustka A.;
RT "Towards a catalog of human genes and proteins: sequencing and
RT analysis of 500 novel complete protein coding human cDNAs.";
RL Genome Res. 11:422-435(2001).
CC -!- FUNCTION: Involved in the maintenance of nuclear organization and
CC structural integrity. Probable anchoring protein which theters the
CC nucleus to the cytoskeleton. Connects nuclei to the cytoskeleton
CC by interacting with the nuclear envelope and with F-actin in the
CC cytoplasm.
CC -!- SUBUNIT: Interacts with F-actin via its N-terminal domain.
CC -!- SUBCELLULAR LOCATION: Type IV membrane protein (Potential). The
CC largest part of the protein is cytoplasmic, while its C-terminal
CC part is associated with the nuclear envelope, most probably the
CC outer nuclear membrane. Remains associated with the nuclear
CC envelope during its breakdown in mitotic cells.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=9;
CC Name=1;
CC IsoId=Q8WXH0-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8WXH0-2; Sequence=VSP_007164, VSP_007166;
CC Note=No experimental confirmation available;
CC Name=3;
CC IsoId=Q8WXH0-3; Sequence=VSP_007155;
CC Note=Produced by exon skipping that results in a frameshift. No
CC experimental confirmation available;
CC Name=4; Synonyms=Beta;
CC IsoId=Q8WXH0-4; Sequence=VSP_007156;
CC Name=5; Synonyms=Alpha;
CC IsoId=Q8WXH0-5; Sequence=VSP_007157, VSP_007164, VSP_007165;
CC Name=6;

```
CC         IsoId=Q8WXH0-6; Sequence=VSP_007158, VSP_007165, VSP_007166;
CC         Note=No experimental confirmation available;
CC       Name=7; Synonyms=Gamma;
CC         IsoId=Q8WXH0-7; Sequence=VSP_007154, VSP_007163;
CC       Name=8;
CC         IsoId=Q8WXH0-8; Sequence=VSP_007161, VSP_007162;
CC         Note=No experimental confirmation available;
CC       Name=9; Synonyms=NUANCE-N-33;
CC         IsoId=Q8WXH0-9; Sequence=VSP_007159, VSP_007160;
CC   -!- TISSUE SPECIFICITY: Widely expressed, with higher level in kidney,
CC       adult and fetal liver, stomach and placenta. Weakly expressed in
CC       skeletal muscle and brain. Isoform 5 is highly expressed in
CC       pancreas, skeletal muscle and heart.
CC   -!- DOMAIN: The Klarsicht domain mediates the nuclear envelope
CC       targeting.
CC   -!- SIMILARITY: Belongs to the nesprin family.
CC   -!- SIMILARITY: Contains 1 actin-binding domain.
CC   -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC   -!- SIMILARITY: Contains 1 Klarsicht domain.
CC   -!- SIMILARITY: Contains 13 leucine-rich (LRR) repeats.
CC   -!- SIMILARITY: Contains 9 spectrin repeats.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF435010; AAL33547.1; -.
DR   EMBL; AF435011; AAL33548.1; -.
DR   EMBL; AY061757; AAL33800.1; -.
DR   EMBL; AY061758; AAL33801.1; -.
DR   EMBL; AY061759; AAL33802.1; -.
DR   EMBL; AF495911; AAN60443.1; -.
DR   EMBL; AL117404; CAB55905.1; -.
DR   EMBL; AL162832; -; NOT_ANNOTATED_CDS.
DR   EMBL; AL355094; -; NOT_ANNOTATED_CDS.

  Query Match             72.1%;  Score 31;  DB 1;  Length 6885;
  Best Local Similarity   50.0%;  Pred. No. 2.3e+03;
  Matches    4;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              |:|:|:|:
Db       6256 SILVWLTE 6263


RESULT 17
SNE1_HUMAN
ID   SNE1_HUMAN     STANDARD;      PRT;  8797 AA.
AC   Q8NF91; O94890; Q8N9P7; Q8TCP1; Q8WWW6; Q8WWW7; Q8WXF6; Q96N17;
AC   Q9C0A7; Q9H525; Q9H526; Q9NS36; Q9NU50; Q9UJ06; Q9UJ07; Q9ULF8;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Nesprin 1 (Nuclear envelope spectrin repeat protein 1) (Synaptic
DE   nuclear envelope protein 1) (Syne-1) (Myocyte nuclear envelope protein
DE   1) (Myne-1) (Enaptin).
GN   SYNE1 OR MYNE1 OR KIAA0796 OR KIAA1756 OR KIAA1262.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), CHARACTERIZATION, AND
RP   MUTAGENESIS OF 8758-LEU--CYS-8763.
RC   TISSUE=Heart, Placenta, Skeletal muscle, Spleen, and Testis;
RX   MEDLINE=21652858; PubMed=11792814;
RA   Zhang Q., Skepper J.N., Yang F., Davies J.D., Hegyi L., Roberts R.G.,
RA   Weissberg P.L., Ellis J.A., Shanahan C.M.;
RT   "Nesprins: a novel family of spectrin-repeat-containing proteins that
RT   localize to the nuclear membrane in multiple tissues.";
RL   J. Cell Sci. 114:4485-4498(2001).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND SUBCELLULAR LOCATION.
RC   TISSUE=Heart, Spleen, and Testis;
RX   MEDLINE=22296983; PubMed=12408964;
RA   Zhang Q., Ragnauth C., Greener M.J., Shanahan C.M., Roberts R.G.;
RT   "The nesprins are giant actin-binding proteins, orthologous to
RT   Drosophila melanogaster muscle protein MSP-300.";
RL   Genomics 80:473-481(2002).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 4), AND VARIANT GLY-8323.
RA   Braune S., Abraham S., Padmakumar V., Tunggal B., Noegel A.A.,
RA   Korenbaum E.;
RT   "The longest isoform of enaptin/Syne-1, a nuclear envelope associated
RT   protein, binds actin cytoskeleton via the alpha-actinin-like actin-
RT   binding domain.";
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORMS 8 AND 9).
RA   Zhang Q., Shanahan C.M.;
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Almeida J., Clark S., Griffiths C., Lloyd D., Parker A., Smith M.,
RA   Tracey A., Williams S.;
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE OF 1-856 FROM N.A.
RC   TISSUE=Kidney;
RA   Gough L., Fan J., Lisa G., Chu S., Winnick S., Beck K.A.;
RT   "Golgi localization of syne-1.";
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE OF 28-778 AND 2901-3476 FROM N.A.
RC   TISSUE=Adrenal gland, and Teratocarcinoma;
RA   Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA   Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA   Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA   Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA   Takemoto M., Ota T., Suzuki Y., Sugano S., Nagahari K., Masuho Y.,
RA   Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE OF 443-8797 FROM N.A. (ISOFORM 5).
RC   TISSUE=Brain;
RX   MEDLINE=21082932; PubMed=11214970;
RA   Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIX.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 7:347-355(2000).
RN   [9]
RP   SEQUENCE OF 743-8797 FROM N.A. (ISOFORM 6).
RC   TISSUE=Brain;
RA   Ansorge W., Wirkner U., Mewes H.-W., Weil B., Wiemann S.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE OF 4219-8797 FROM N.A. (ISOFORM 7).
RC   TISSUE=Brain;
RX   MEDLINE=20039619; PubMed=10574462;
RA   Nagase T., Ishikawa K.-I., Kikuno R., Hirosawa M., Nomura N.,
RA   Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XV.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:337-345(1999).
RN   [11]
RP   SEQUENCE OF 6922-8797 FROM N.A.
```

```
RC   TISSUE=Brain;
RX   MEDLINE=99087487; PubMed=9872452;
RA   Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XI.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 5:277-286(1998).
RN   [12]
RP   REVISIONS.
RC   TISSUE=Brain;
RX   MEDLINE=22158633; PubMed=12168954;
RA   Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT   "Construction of expression-ready cDNA clones for KIAA genes: manual
RT   curation of 330 KIAA cDNA clones.";
RL   DNA Res. 9:99-106(2002).
RN   [13]
RP   SEQUENCE OF 7631-8797 FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=21659781; PubMed=11801724;
RA   Mislow J.M.K., Kim M.S., Davis D.B., McNally E.M.;
RT   "Myne-1, a spectrin repeat transmembrane protein of the myocyte inner
RT   nuclear membrane, interacts with lamin A/C.";
RL   J. Cell Sci. 115:61-70(2002).
RN   [14]
RP   SEQUENCE OF 8406-8797 FROM N.A.
RA   Ma F.-R., Zhu L.-P.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Involved in the maintenance of nuclear organization and
CC       structural integrity. Probable anchoring protein which theters the
CC       nucleus to the cytoskeleton. Connects nuclei to the cytoskeleton
CC       by interacting with the nuclear envelope and with F-actin in the
CC       cytoplasm.
CC   -!- SUBUNIT: Interacts with MUSK, with F-actin via its N-terminal
CC       domain, and with LMNA in vitro (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type IV membrane protein (Potential). The
CC       largest part of the protein is cytoplasmic, while its C-terminal
CC       part is associated with the nuclear envelope, most probably the
CC       outer nuclear membrane. In skeletal and smooth muscles, a
CC       significant amount is found in the sarcomeres.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=9;
CC       Name=1;
CC         IsoId=Q8NF91-1; Sequence=Displayed;
CC       Name=2; Synonyms=Beta;
CC         IsoId=Q8NF91-2; Sequence=VSP_007130;
CC       Name=3; Synonyms=Alpha;
CC         IsoId=Q8NF91-3; Sequence=VSP_007132, VSP_007144;
CC       Name=4;
CC         IsoId=Q8NF91-4; Sequence=VSP_007134, VSP_007139, VSP_007140,
CC                                  VSP_007144;
CC       Name=5;
CC         IsoId=Q8NF91-5; Sequence=VSP_007135, VSP_007136;
CC         Note=No experimental confirmation available;
CC       Name=6;
CC         IsoId=Q8NF91-6; Sequence=VSP_007137, VSP_007138;
CC         Note=No experimental confirmation available;
CC       Name=7;
CC         IsoId=Q8NF91-7; Sequence=VSP_007141, VSP_007142;
CC         Note=No experimental confirmation available;
CC       Name=8; Synonyms=Beta 2;
CC         IsoId=Q8NF91-8; Sequence=VSP_007131;
CC       Name=9; Synonyms=Alpha 2;
CC         IsoId=Q8NF91-9; Sequence=VSP_007133, VSP_007143, VSP_007144;
CC   -!- TISSUE SPECIFICITY: Widely expressed. Highly expressed in skeletal
CC       and smooth muscles, heart, spleen, and peripheral blood
CC       leukocytes.
CC   -!- DOMAIN: The Klarsicht domain, which contains a transmembrane
CC       domain, mediates the nuclear envelope targeting.
CC   -!- SIMILARITY: Belongs to the nesprin family.
CC   -!- SIMILARITY: Contains 1 actin-binding domain.
CC   -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC   -!- SIMILARITY: Contains 12 HAT repeats.
CC   -!- SIMILARITY: Contains 1 Klarsicht domain
CC   -!- SIMILARITY: Contains 31 spectrin repeats.
CC   -!- CAUTION: Ref.5 (CAB55865, CAB55866, CAC16280.and CAC16281)
CC       sequences differ from that shown due to erroneous gene model
CC       prediction.
CC   -!- CAUTION: Ref.7 (BAB71097) sequence differs from that shown due to
CC       a chimeric cDNA.
CC   -!- CAUTION: Ref.14 sequence differs from that shown due to two
CC       frameshifts in positions 8412 and 8784.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AY061755; AAL33798.1; -.
DR   EMBL; AY061756; AAL33799.1; -.
DR   EMBL; AF495910; AAN60442.1; -.
DR   EMBL; AF535142; AAN03486.1; -.
DR   EMBL; AY184203; AAO27771.1; -.
DR   EMBL; AY184206; AAO27774.1; -.
DR   EMBL; AL049548; CAB55865.1; ALT_SEQ.
DR   EMBL; AL049548; CAB55866.1; -.
DR   EMBL; AL078582; CAB87586.1; -.
DR   EMBL; AL136079; -; NOT_ANNOTATED_CDS.
DR   EMBL; AL138832; CAC16280.1; ALT_SEQ.
DR   EMBL; AL138832; CAC16281.1; ALT_SEQ.
DR   EMBL; AL357081; -; NOT_ANNOTATED_CDS.
DR   EMBL; AL450401; -; NOT_ANNOTATED_CDS.
DR   EMBL; AL589963; -; NOT_ANNOTATED_CDS.

  Query Match             72.1%;  Score 31;  DB 1;  Length 8797;
  Best Local Similarity   50.0%;  Pred. No. 3e+03;
  Matches    4;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              |:|:|:|:
Db       8119 SILVWLTE 8126


RESULT 18
VE5_HPV16
ID   VE5_HPV16      STANDARD;      PRT;    83 AA.
AC   P06927;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Probable E5 protein.
GN   E5.
OS   Human papillomavirus type 16.
OC   Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC   Papillomavirus.
OX   NCBI_TaxID=10581;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88160059; PubMed=2831662;
RA   Bubb V., McCance D.J., Schlegel R.;
RT   "DNA sequence of the HPV-16 E5 ORF and the structural conservation of
RT   its encoded protein.";
RL   Virology 163:243-246(1988).
RN   [2]
RP   SEQUENCE OF 20-83 FROM N.A.
RX   MEDLINE=85246220; PubMed=2990099;
RA   Seedorf K., Krammer G., Durst M., Suhai S., Rowekamp W.G.;
RT   "Human papillomavirus type 16 DNA sequence.";
RL   Virology 145:181-185(1985).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; K02718; AAA46938.1; ALT_SEQ.
DR   PIR; A30016; W5WLHS.
DR   InterPro; IPR004270; Papilloma_E5.
DR   Pfam; PF03025; Papilloma_E5; 1.
KW   Early protein.
SQ   SEQUENCE   83 AA;  9401 MW;  442C0ABF0D77CDCF CRC64;

  Query Match             69.8%;  Score 30;  DB 1;  Length 83;
  Best Local Similarity   83.3%;  Pred. No. 46;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LLMWIT 7
              ||:|||
Db         47 LLLWIT 52


RESULT 19
KV40_HUMAN
ID   KV40_HUMAN     STANDARD;      PRT;   121 AA.
AC   P06312;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Ig kappa chain V-IV region precursor (Fragment).
GN   IGKV4-1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86041853; PubMed=2997712;
RA   Klobeck H.G., Bornkamm G.W., Combriato G., Mocikat R., Pohlenz H.D.,
RA   Zachau H.G.;
RT   "Subgroup IV of human immunoglobulin K light chains is encoded by a
RT   single germline gene.";
RL   Nucleic Acids Res. 13:6515-6529(1985).
CC   -!- MISCELLANEOUS: THERE IS ONLY ONE IG KAPPA V-IV GENE.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z00023; CAA77318.1; -.
DR   PIR; A01902; K4HU.
DR   HSSP; P80362; 1WTL.
DR   Genew; HGNC:5834; IGKV4-1.
DR   GO; GO:0005576; C:extracellular; NAS.
DR   GO; GO:0003823; F:antigen binding; NAS.
DR   GO; GO:0006955; P:immune response; NAS.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Immunoglobulin V region; Signal.
FT   SIGNAL        1     20
FT   CHAIN        21   >121       IG KAPPA CHAIN V-IV REGION.
FT   DOMAIN       21     43       FRAMEWORK-1.
FT   DOMAIN       44     60       COMPLEMENTARITY-DETERMINING-1.
FT   DOMAIN       61     75       FRAMEWORK-2.
FT   DOMAIN       76     82       COMPLEMENTARITY-DETERMINING-2.
FT   DOMAIN       83    114       FRAMEWORK-3.
FT   DOMAIN      115    121       COMPLEMENTARITY-DETERMINING-3.
FT   DISULFID     43    114       BY SIMILARITY.
FT   NON_TER     121    121
SQ   SEQUENCE   121 AA;  13380 MW;  9586AD4188D33974 CRC64;

  Query Match             69.8%;  Score 30;  DB 1;  Length 121;
  Best Local Similarity   71.4%;  Pred. No. 67;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWIT 7
              |||:||:
Db         10 SLLLWIS 16


RESULT 20
KV4B_HUMAN
ID   KV4B_HUMAN     STANDARD;      PRT;   133 AA.
AC   P06313;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Ig kappa chain V-IV region JI precursor.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86041853; PubMed=2997712;
RA   Klobeck H.G., Bornkamm G.W., Combriato G., Mocikat R., Pohlenz H.D.,
RA   Zachau H.G.;
RT   "Subgroup IV of human immunoglobulin K light chains is encoded by a
RT   single germline gene.";
RL   Nucleic Acids Res. 13:6515-6529(1985).
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z00022; CAA77317.1; -.
DR   PIR; A01904; K4HUJI.
DR   HSSP; P80362; 1WTL.
DR   GO; GO:0005576; C:extracellular; NAS.
DR   GO; GO:0003823; F:antigen binding; NAS.
DR   GO; GO:0006955; P:immune response; NAS.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Immunoglobulin V region; Signal.
FT   SIGNAL        1     20
FT   CHAIN        21    133       IG KAPPA CHAIN V-IV REGION JI.
FT   DOMAIN       21     43       FRAMEWORK-1.
FT   DOMAIN       44     60       COMPLEMENTARITY-DETERMINING-1.
FT   DOMAIN       61     75       FRAMEWORK-2.
FT   DOMAIN       76     82       COMPLEMENTARITY-DETERMINING-2.
FT   DOMAIN       83    114       FRAMEWORK-3.
FT   DOMAIN      115    122       COMPLEMENTARITY-DETERMINING-3.
FT   DOMAIN      123    132       FRAMEWORK-4.
FT   DISULFID     43    114       BY SIMILARITY.
FT   NON_TER     133    133
SQ   SEQUENCE   133 AA;  14632 MW;  5FB3953066744AF4 CRC64;

  Query Match             69.8%;  Score 30;  DB 1;  Length 133;
  Best Local Similarity   71.4%;  Pred. No. 73;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWIT 7
              |||:||:
Db         10 SLLLWIS 16
```

```
RESULT 21
KV4C_HUMAN
ID   KV4C_HUMAN     STANDARD;      PRT;   134 AA.
AC   P06314;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-APR-1988 (Rel. 07, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Ig kappa chain V-IV region B17 precursor.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86041854; PubMed=2997713;
RA   Marsh P., Mills F., Gould H.;
RT   "Detection of a unique human V kappa IV germline gene by a cloned
RT   cDNA probe.";
RL   Nucleic Acids Res. 13:6531-6544(1985).
RN   [2]
RP   REVISION TO 76.
RA   Marsh P.;
RL   Submitted (OCT-1986) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X02990; CAA26733.1; -.
DR   HSSP; P80362; 1WTL.
DR   GO; GO:0005576; C:extracellular; NAS.
DR   GO; GO:0003823; F:antigen binding; NAS.
DR   GO; GO:0006955; P:immune response; NAS.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Immunoglobulin V region; Signal.
FT   SIGNAL        1     20
FT   CHAIN        21    134       IG KAPPA CHAIN V-IV REGION B17.
FT   DOMAIN       21     43       FRAMEWORK-1.
FT   DOMAIN       44     60       COMPLEMENTARITY-DETERMINING-1.
FT   DOMAIN       61     75       FRAMEWORK-2.
FT   DOMAIN       76     82       COMPLEMENTARITY-DETERMINING-2.
FT   DOMAIN       83    114       FRAMEWORK-3.
FT   DOMAIN      115    121       COMPLEMENTARITY-DETERMINING-3.
FT   DOMAIN      122    133       FRAMEWORK-4.
FT   DISULFID     43    114       BY SIMILARITY.
FT   NON_TER     134    134
SQ   SEQUENCE   134 AA;  14966 MW;  6413A22FD0738832 CRC64;

  Query Match             69.8%;  Score 30;  DB 1;  Length 134;
  Best Local Similarity   71.4%;  Pred. No. 74;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWIT 7
              |||:||:
Db         10 SLLLWIS 16


RESULT 22
HRPX_PLALO
ID   HRPX_PLALO     STANDARD;      PRT;   351 AA.
AC   P04929;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Histidine-rich glycoprotein precursor.
OS   Plasmodium lophurae.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5853;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85061618; PubMed=6095114;
RA   Ravetch J.V., Feder R., Pavlovec A., Blobel G.;
RT   "Primary structure and genomic organization of the histidine-rich
RT   protein of the malaria parasite Plasmodium lophurae.";
RL   Nature 312:616-620(1984).
CC   -!- MISCELLANEOUS: In the intraerythrocytic stages of development of
CC       P.lophurae in ducks, there is a synthesis of a major protein that
CC       accumulates to comprise at least 50% of the cellular mass: the
CC       histidine rich protein.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X01469; CAA25698.1; -.
DR   PIR; A22692; KGZQHL.
KW   Malaria; Repeat; Glycoprotein; Signal.
FT   SIGNAL        1     23
FT   PROPEP       24     47
FT   CHAIN        48    351       HISTIDINE-RICH GLYCOPROTEIN.
FT   CARBOHYD     40     40       N-LINKED (GLCNAC. . .) (PROBABLE).
FT   DOMAIN       59     90       2 X 16 AA TANDEM REPEATS.
FT   REPEAT       59     74       16-1.
FT   REPEAT       75     90       16-2.
FT   DOMAIN       91    123       2 X 17 AA TANDEM REPEATS.
FT   REPEAT       91    107       17-1.
FT   REPEAT      108    123       17-2.
FT   DOMAIN      124    153       2 X 15 AA TANDEM REPEATS.
FT   REPEAT      124    138       15-1.
FT   REPEAT      139    153       15-2.
FT   DOMAIN      173    351       18 X 10 AA TANDEM REPEATS.
SQ   SEQUENCE   351 AA;  44032 MW;  D19A48D47D890453 CRC64;

  Query Match             69.8%;  Score 30;  DB 1;  Length 351;
  Best Local Similarity   62.5%;  Pred. No. 1.9e+02;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              | |:||:|
Db         12 SFLVWISQ 19


RESULT 23
HOFC_HAEIN
ID   HOFC_HAEIN     STANDARD;      PRT;   406 AA.
AC   P44621;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Protein transport protein hofC homolog.
GN   HOFC OR HOPC OR HI0297.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
```

```
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Probable).
CC   -!- SIMILARITY: BELONGS TO THE PULF/OUTF/EXEF/XPSF/XCPS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32715; AAC21961.1; -.
DR   PIR; D64060; D64060.
DR   TIGR; HI0297; -.
DR   InterPro; IPR003004; Bac_GSPF.
DR   InterPro; IPR001992; Bact_secr_systII.
DR   Pfam; PF00482; GSPII_F; 1.
DR   PRINTS; PR00812; BCTERIALGSPF.
DR   PROSITE; PS00874; T2SP_F; 1.
KW   Transport; Transmembrane; Inner membrane; Complete proteome.
FT   TRANSMEM    167    187       POTENTIAL.
FT   TRANSMEM    214    234       POTENTIAL.
FT   TRANSMEM    379    399       POTENTIAL.
SQ   SEQUENCE   406 AA;  46290 MW;  429D83B7C10F8F82 CRC64;

  Query Match           69.8%;  Score 30;  DB 1;  Length 406;
  Best Local Similarity   50.0%;  Pred. No. 2.2e+02;
  Matches    4;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              |:| |::|
Db        312 SILQWVSQ 319


RESULT 24
RGS9_MOUSE
ID   RGS9_MOUSE     STANDARD;      PRT;   675 AA.
AC   O54828; Q9Z0S0;
DT   15-DEC-1998 (Rel. 37, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Regulator of G-protein signaling 9 (RGS9).
GN   RGS9.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=98119533; PubMed=9459445;
RA   He W., Cowan C.W., Wensel T.G.;
RT   "RGS9, a GTPase accelerator for phototransduction.";
RL   Neuron 20:95-102(1998).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   STRAIN=C57BL/6; TISSUE=Forebrain;
RX   MEDLINE=99165807; PubMed=10066255;
RA   Rahman Z., Gold S.J., Potenza M.N., Cowan C.W., Ni Y.G., He W.,
RA   Wensel T.G., Nestler E.J.;
RT   "Cloning and characterization of RGS9-2: a striatal-enriched
RT   alternatively spliced product of the RGS9 gene.";
RL   J. Neurosci. 19:2016-2026(1999).
RN   [3]
RP   PHOSPHORYLATION OF ISOFORM 1.
RX   MEDLINE=21303582; PubMed=11292825;
RA   Hu G., Jang G.F., Cowan C.W., Wensel T.G., Palczewski K.;
RT   "Phosphorylation of RGS9-1 by an endogenous protein kinase in rod
RT   outer segments.";
RL   J. Biol. Chem. 276:22287-22295(2001).
CC   -!- FUNCTION: INHIBITS SIGNAL TRANSDUCTION BY INCREASING THE GTPASE
CC       ACTIVITY OF G PROTEIN ALPHA SUBUNITS THEREBY DRIVING THEM INTO
CC       THEIR INACTIVE GDP-BOUND FORM. BINDS TO G(T)-ALPHA. INVOLVED IN
CC       PHOTOTRANSDUCTION; KEY ELEMENT IN THE RECOVERY PHASE OF VISUAL
CC       TRANSDUCTION.
CC   -!- SUBUNIT: HETERODIMER WITH GBETA5 (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=2;
CC         IsoId=O54828-1; Sequence=Displayed;
CC       Name=1;
CC         IsoId=O54828-2; Sequence=VSP_005678, VSP_005679;
CC   -!- TISSUE SPECIFICITY: ISOFORM 1 IS EXPRESSED IN PHOTORECEPTOR OUTER
CC       SEGMENTS. ISOFORM 2 IS EXPRESSED IN BRAIN STRIATUM.
CC   -!- PTM: RETINAL ISOFORM 1 IS LIGHT-DEPENDENT PHOSPHORYLATED AT SER-
CC       475. PHOSPHORYLATION IS DECREASED BY LIGHT EXPOSITION.
CC   -!- SIMILARITY: Contains 1 RGS domain.
CC   -!- SIMILARITY: Contains 1 G protein gamma domain.
CC   -!- SIMILARITY: Contains 1 DEP domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF011358; AAC99481.1; -.
DR   EMBL; AF125046; AAD20014.1; -.
DR   HSSP; P49795; 1CMZ.
DR   MGD; MGI:1338824; Rgs9.
DR   InterPro; IPR000591; DEP.
DR   InterPro; IPR001770; G-gamma.
DR   InterPro; IPR000342; Regl_Gprotein.
DR   Pfam; PF00610; DEP; 1.
DR   Pfam; PF00631; G-gamma; 1.
DR   Pfam; PF00615; RGS; 1.
DR   PRINTS; PR01301; RGSPROTEIN.
DR   ProDom; PD001580; Regl_Gprotein; 1.
DR   SMART; SM00049; DEP; 1.
DR   SMART; SM00224; GGL; 1.
DR   SMART; SM00315; RGS; 1.
DR   PROSITE; PS50186; DEP; 1.
DR   PROSITE; PS50058; G_PROTEIN_GAMMA; FALSE_NEG.
DR   PROSITE; PS50132; RGS; 1.
KW   Signal transduction inhibitor; Alternative splicing; Phosphorylation.
FT   DOMAIN       30    105       DEP.
FT   DOMAIN      222    283       G PROTEIN GAMMA-LIKE.
FT   DOMAIN      299    414       RGS.
FT   VARSPLIC    467    484       PGQHLAPSPHLAVYTGTC -> VMSKLDRRSQLKKELPPK
FT                                (in isoform 1).
FT                                /FTId=VSP_005678.
FT   VARSPLIC    485    675       Missing (in isoform 1).
FT                                /FTId=VSP_005679.
SQ   SEQUENCE   675 AA;  76971 MW;  0EC910D833FFD06D CRC64;

  Query Match           69.8%;  Score 30;  DB 1;  Length 675;
  Best Local Similarity   71.4%;  Pred. No. 3.6e+02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LLMWITQ 8
              :| ||||
Db         55 VLQWITQ 61


RESULT 25
```

```
KALM_CHICK
ID   KALM_CHICK     STANDARD;      PRT;   676 AA.
AC   P33005;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Anosmin 1 precursor (Kallmann syndrome protein homolog).
GN   KAL.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=94010957; PubMed=8406507;
RA   Legouis R., Cohen-Salmon M., del Castillo I., Levilliers J.,
RA   Capy L., Mornon J.-P., Petit C.;
RT   "Characterization of the chicken and quail homologues of the human
RT   gene responsible for the X-linked Kallmann syndrome.";
RL   Genomics 17:516-518(1993).
CC   -!- FUNCTION: May be an adhesion-like molecule with anti-protease
CC       activity.
CC   -!- SUBCELLULAR LOCATION: Secreted. Localized at cell surface (By
CC       similarity).
CC   -!- SIMILARITY: Contains 4 fibronectin type III domains.
CC   -!- SIMILARITY: Contains 1 WAP-type domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L12144; AAA51435.1; -.
DR   PIR; B47222; B47222.
DR   HSSP; P19957; 2REL.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008197; WAP.
DR   Pfam; PF00041; fn3; 3.
DR   Pfam; PF00095; wap; 1.
DR   PRINTS; PR00003; 4DISULPHCORE.
DR   SMART; SM00060; FN3; 3.
DR   SMART; SM00217; WAP; 1.
DR   PROSITE; PS00317; 4_DISULFIDE_CORE; 1.
KW   Cell adhesion; Glycoprotein; Serine protease inhibitor; Repeat;
KW   Signal.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    676       ANOSMIN 1.
FT   DOMAIN       22    115       "CYSTEINE BOX".
FT   DOMAIN      125    171       WAP.
FT   DOMAIN      176    280       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      281    397       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      398    536       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      537    657       FIBRONECTIN TYPE-III 4.
FT   CARBOHYD     66     66       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    204    204       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    295    295       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    466    466       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    549    549       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    560    560       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   676 AA;  76375 MW;  3FAC7ED82EA7E352 CRC64;

  Query Match             69.8%;  Score 30;  DB 1;  Length 676;
  Best Local Similarity   57.1%;  Pred. No. 3.6e+02;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWIT 7
              :||:|:|
Db         12 ALLLWVT 18


Search completed: August 22, 2004, 11:00:25
Job time : 24 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 22, 2004, 10:55:43 ; Search time 39 Seconds
(without alignments)
22.198 Million cell updates/sec

Title: US-10-706-475-10
Perfect score: 43
Sequence: 1 SLLMWITQX 9

Scoring table: BLOSUM62XX
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	36	83.7	196	2	AE2876	conserved hypothet
2	36	83.7	198	2	G97652	hypothetical prote
3	35	81.4	861	2	S77086	hypothetical prote
4	34	79.1	122	2	S54492	probable membrane
5	34	79.1	299	2	D84519	probable endoxylog
6	33	76.7	479	2	F84179	hypothetical prote
7	32	74.4	145	2	A64469	hypothetical prote
8	32	74.4	278	2	G81139	hypothetical prote
9	32	74.4	456	1	I40516	spaF protein - Bac
10	32	74.4	724	2	A38748	3-phosphatidylinos
11	32	74.4	724	2	A38749	3-phosphatidylinos
12	32	74.4	724	2	A38747	phosphatidlyinosit
13	32	74.4	999	2	AG2413	hypothetical prote
14	32	74.4	1076	2	B96682	protein F1E22.14 [
15	32	74.4	2330	1	RRIWMV	genome polyprotein
16	32	74.4	2331	2	S44054	genome polyprotein
17	31	72.1	151	2	E69786	ribosomal-protein-
18	31	72.1	213	2	T07603	NADH2 dehydrogenas
19	31	72.1	228	2	S45677	proteinase inhibit
20	31	72.1	311	2	T11362	NADH2 dehydrogenas
21	31	72.1	349	2	B83434	translocation prot
22	31	72.1	437	2	C49274	protein farnesyltr
23	31	72.1	437	2	B49274	protein farnesyltr
24	31	72.1	437	2	A40037	protein farnesyltr
25	31	72.1	752	2	H96603	unknown protein F1
26	31	72.1	897	2	G89923	hypothetical prote
27	31	72.1	1092	2	T12520	hypothetical prote
28	31	72.1	1398	2	T20434	hypothetical prote
29	30	69.8	53	2	H82027	hypothetical prote
30	30	69.8	83	1	W5WLHS	E5 protein - human
31	30	69.8	121	1	K4HU	Ig kappa chain pre
32	30	69.8	129	2	S40347	Ig kappa chain - h
33	30	69.8	132	2	S46373	Ig kappa chain V-J
34	30	69.8	133	1	K4HUJI	Ig kappa chain pre
35	30	69.8	134	1	K4HU17	Ig kappa chain pre
36	30	69.8	134	2	S21917	Ig kappa chain V r
37	30	69.8	134	2	S49531	anti-Sm antibody V
38	30	69.8	135	2	F86756	prophage pi2 prote
39	30	69.8	136	2	A49137	Ig kappa chain pre
40	30	69.8	138	2	A53261	Ig kappa chain pre
41	30	69.8	168	2	S77282	hypothetical prote
42	30	69.8	170	2	T00221	type II secretion
43	30	69.8	192	2	D90464	conserved hypothet
44	30	69.8	231	2	E64711	conserved hypothet
45	30	69.8	237	2	D83407	probable transcrip
46	30	69.8	240	2	S06084	Ig kappa chain pre
47	30	69.8	247	1	E69021	conserved hypothet
48	30	69.8	262	2	H64623	spermidine synthas
49	30	69.8	262	2	C71891	probable spermidin
50	30	69.8	281	2	T29825	hypothetical prote

ALIGNMENTS

RESULT 1
AE2876
conserved hypothetical protein Atu2439 [imported] - Agrobacterium tumefaciens (strain C
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Dec-2002
C;Accession: AE2876
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClel
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AE2876
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-196 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL43427.1; PID:g17740928; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu2439
A;Map position: circular chromosome
C;Superfamily: uncharacterized conserved protein

Query Match 83.7%; Score 36; DB 2; Length 196;
Best Local Similarity 85.7%; Pred. No. 15;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 SLLMWIT 7
              |||||:|
Db         23 SLLMWVT 29
```

RESULT 2
G97652
hypothetical protein AGR_C_4424 [imported] - Agrobacterium tumefaciens (strain C58, Cer
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Dec-2002
C;Accession: G97652
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tu
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: G97652
A;Status: preliminary

A;Molecule type: DNA
A;Residues: 1-198 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK88176.1; PID:g15157620; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_4424
A;Map position: circular chromosome
C;Superfamily: uncharacterized conserved protein

Query Match 83.7%; Score 36; DB 2; Length 198;
Best Local Similarity 85.7%; Pred. No. 15;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLMWIT 7
|||||:|
Db 25 SLLMWVT 31

RESULT 3
S77086
hypothetical protein sll0737 - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C;Accession: S77086
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S77086
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-861 <KAN>
A;Cross-references: EMBL:D64005; GB:AB001339; NID:g1001779; PIDN:BAA10778.1; PID:g100662
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Superfamily: Synechocystis hypothetical protein sll0737

Query Match 81.4%; Score 35; DB 2; Length 861;
Best Local Similarity 75.0%; Pred. No. 95;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SLLMWITQ 8
|||:| ||
Db 157 SLLLWFTQ 164

RESULT 4
S54492
probable membrane protein YMR123w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YM8564.05
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jul-1995 #sequence_revision 17-Nov-1995 #text_change 19-Apr-2002
C;Accession: S54492
R;Lye, G.; Churcher, C.M.
submitted to the EMBL Data Library, May 1995
A;Reference number: S54014
A;Accession: S54492
A;Molecule type: DNA
A;Residues: 1-122 <LYE>
A;Cross-references: EMBL:Z49273; NID:g809577; PID:g809582; GSPDB:GN00013; MIPS:YMR123w
C;Genetics:
A;Gene: SGD:PKR1; MIPS:YMR123w
A;Cross-references: SGD:S0004730
A;Map position: 13R
C;Superfamily: Saccharomyces cerevisiae probable membrane protein YMR123w
C;Keywords: transmembrane protein
F;21-37/Domain: transmembrane #status predicted <TM1>
F;51-67/Domain: transmembrane #status predicted <TM2>

Query Match 79.1%; Score 34; DB 2; Length 122;
Best Local Similarity 85.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLMWIT 7
|||:|||
Db 56 SLLLWIT 62

RESULT 5
D84519
probable endoxyloglucan glycosyltransferase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: D84519
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: D84519
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-299 <STO>
A;Cross-references: GB:AE002093; NID:g3810598; PIDN:AAC69380.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g14620
A;Map position: 2
C;Superfamily: endoxyloglucan transferase

Query Match 79.1%; Score 34; DB 2; Length 299;
Best Local Similarity 62.5%; Pred. No. 52;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLMWITQ 8
|||:|::|
Db 21 SLLLWVSQ 28

RESULT 6
F84179
hypothetical protein Vng0189c [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: F84179
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: F84179
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-479 <STO>
A;Cross-references: GB:AE004437; NID:g10579836; PIDN:AAG18802.1; GSPDB:GN00138
C;Genetics:
A;Gene: VNG0189C

Query Match 76.7%; Score 33; DB 2; Length 479;
Best Local Similarity 62.5%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SLLMWITQ 8
: :| ||:||
Db 263 ALYMWVTQ 270

RESULT 7
A64469
hypothetical protein MJ1354 - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: A64469

R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: A64469
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-145 <BUL>
A;Cross-references: GB:U67575; GB:L77117; NID:g1591992; PIDN:AAB99366.1; PID:g1591997; T
C;Genetics:
A;Map position: FOR1304340-1304777

  Query Match 74.4%; Score 32; DB 2; Length 145;
  Best Local Similarity 71.4%; Pred. No. 60;
  Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLMWIT 7
  |||:|:|
Db 106 SLLVWVT 112

RESULT 8
G81139
hypothetical protein NMB0938 [imported] - Neisseria meningitidis (strain MC58 serogroup
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: G81139
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: G81139
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-278 <TET>
A;Cross-references: GB:AE002445; GB:AE002098; NID:g7226173; PIDN:AAF41344.1; PID:g722617
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB0938

  Query Match 74.4%; Score 32; DB 2; Length 278;
  Best Local Similarity 71.4%; Pred. No. 1.1e+02;
  Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLMWIT 7
  |||:|:|
Db 7 SLLLWLT 13

RESULT 9
I40516
spaF protein - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
C;Accession: I40516
R;Klein, C.; Entian, K.D.
Appl. Environ. Microbiol. 60, 2793-2801, 1994
A;Title: Genes involved in self-protection against the lantibiotic subtilin produced by
A;Reference number: I40511; MUID:94368094; PMID:8085823
A;Accession: I40516
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-456 <RES>
A;Cross-references: EMBL:U09819; NID:g595315; PID:g595321
C;Genetics:
A;Gene: spaF
C;Superfamily: Bacillus subtilis spaF protein; ATP-binding cassette homology
C;Keywords: ATP; nucleotide binding; P-loop
F;30-214/Domain: ATP-binding cassette homology <ABC>
F;47-54/Region: nucleotide-binding motif A (P-loop)

  Query Match 74.4%; Score 32; DB 1; Length 456;
  Best Local Similarity 71.4%; Pred. No. 1.8e+02;
  Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLMWIT 7
  |||:|:|
Db 378 SLLLWLT 384

RESULT 10
A38748
3-phosphatidylinositol kinase (EC 2.7.1.-) 85K chain - human
C;Species: Homo sapiens (man)
C;Date: 24-Jan-1992 #sequence_revision 24-Jan-1992 #text_change 16-Jul-1999
C;Accession: A38748; S28402
R;Skolnik, E.Y.; Margolis, B.; Mohammadi, M.; Lowenstein, E.; Fischer, R.; Drepps, A.; U
Cell 65, 83-90, 1991
A;Title: Cloning of P13 kinase-associated p85 utilizing a novel method for expression/cl
A;Reference number: A38748; MUID:91191565; PMID:1849461
A;Accession: A38748
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-724 <SKO>
A;Cross-references: GB:M61906
R;Panayotou, G.; Bax, B.; Gout, I.; Federwisch, M.; Wroblowski, B.; Dhand, R.; Fry, M.J
EMBO J. 11, 4261-4272, 1992
A;Title: Interaction of the p85 subunit of PI 3-kinase and its N-terminal SH2 domain wi
A;Reference number: S28402; MUID:93049176; PMID:1330535
A;Accession: S28402
A;Status: preliminary
A;Molecule type: protein
A;Residues: 301-311,315-319;424-439 <PAN>
C;Genetics:
A;Gene: GDB:PIK3R1
A;Cross-references: GDB:127604; OMIM:171833
A;Map position: 5q12-5q13
C;Superfamily: SH2 homology
C;Keywords: phosphotransferase
F;333-428/Domain: SH2 homology <SH2A>
F;624-718/Domain: SH2 homology <SH2>

  Query Match 74.4%; Score 32; DB 2; Length 724;
  Best Local Similarity 83.3%; Pred. No. 2.9e+02;
  Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 LMWITQ 8
  |||:||
Db 581 LMWLTQ 586

RESULT 11
A38749
3-phosphatidylinositol kinase (EC 2.7.1.-) 85K chain alpha - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 14-Feb-1992 #sequence_revision 14-Feb-1992 #text_change 05-Nov-1999
C;Accession: A38749
R;Otsu, M.; Hiles, I.; Gout, I.; Fry, M.J.; Ruiz-Larrea, F.; Panayotou, G.; Thompson, A

Cell 65, 91-104, 1991
A;Title: Characterization of two 85 kd proteins that associate with receptor tyrosine k
A;Reference number: A38749; MUID:91191567; PMID:1707345
A;Accession: A38749
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-724 <OTS>
A;Cross-references: GB:M61746; GB:M61745; NID:g163476; PIDN:AAA79511.1; PID:g163477
C;Superfamily: SH2 homology
C;Keywords: phosphotransferase
F;333-428/Domain: SH2 homology <SH2A>

F;624-718/Domain: SH2 homology <SH2>

Query Match 74.4%; Score 32; DB 2; Length 724;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          3 LMWITQ 8
              |||:||
Db        581 LMWLTQ 586
```

RESULT 12
A38747
phosphatidlyinositol 3-kinase (EC 2.7.1.-) 85K chain - mouse
C;Species: Mus musculus (house mouse)
C;Date: 03-Aug-1992 #sequence_revision 03-Aug-1992 #text_change 16-Jul-1999
C;Accession: A38747
R;Escobedo, J.A.; Navankasattusas, S.; Kavanaugh, W.M.; Milfay, D.; Fried, V.A.; William
Cell 65, 75-82, 1991
A;Title: cDNA cloning of a novel 85 kd protein that has SH2 domains and regulates bindin
A;Reference number: A38747; MUID:91191564; PMID:1849460
A;Accession: A38747
A;Molecule type: mRNA
A;Residues: 1-724 <ESC>
A;Cross-references: GB:M60651
C;Comment: This protein binds a phosphotyrosine-containing sequence of ligand-activated
phosphatidylinositol at position 3 of the inositol ring.
C;Superfamily: SH2 homology
C;Keywords: phosphotransferase
F;333-428/Domain: SH2 homology <SH2A>
F;624-718/Domain: SH2 homology <SH2>

Query Match 74.4%; Score 32; DB 2; Length 724;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          3 LMWITQ 8
              |||:||
Db        581 LMWLTQ 586
```

RESULT 13
AG2413
hypothetical protein alr4863 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AG2413
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AG2413
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-999 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB76562.1; PID:g17134000; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr4863

Query Match 74.4%; Score 32; DB 2; Length 999;
Best Local Similarity 62.5%; Pred. No. 3.9e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 SLLMWITQ 8
              ||| |::|
Db         11 SLLQWVSQ 18
```

RESULT 14
B96682
protein F1E22.14 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: B96682
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: B96682
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1076 <STO>
A;Cross-references: GB:AE005173; NID:g6686403; PIDN:AAF23837.1; GSPDB:GN00141
C;Genetics:
A;Gene: F1E22.14
A;Map position: 1

Query Match 74.4%; Score 32; DB 2; Length 1076;
Best Local Similarity 71.4%; Pred. No. 4.2e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy          2 LLMWITQ 8
              :|:||||
Db       1016 ILVWITQ 1022
```

RESULT 15
RRIWMV
genome polyprotein - Marburg virus (strain Musoke)
N;Alternate names: L protein
N;Contains: RNA-directed RNA polymerase (EC 2.7.7.48)
C;Species: Marburg virus
A;Note: host Homo sapiens (man)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 11-Jun-1999
C;Accession: A42450
R;Muehlberger, E.; Sanchez, A.; Randolf, A.; Will, C.; Kiley, M.P.; Klenk, H.D.; Feldman
Virology 187, 534-547, 1992
A;Title: The nucleotide sequence of the L gene of Marburg virus, a filovirus: homologies
A;Reference number: A42450; MUID:92188528; PMID:1546452
A;Accession: A42450
A;Molecule type: genomic RNA
A;Residues: 1-2330 <MUE>
A;Cross-references: GB:M92834; NID:g332178; PIDN:AAA46562.1; PID:g332179
C;Genetics:
A;Gene: L
C;Superfamily: parainfluenza virus RNA-directed RNA polymerase
C;Keywords: ATP; nucleotidyltransferase; RNA biosynthesis
F;1325-1360/Domain: ATP binding #status predicted <AT1>
F;1390-1420/Domain: ATP binding #status predicted <AT2>
F;1560-1593/Domain: ATP binding #status predicted <AT3>
F;1932-1961/Domain: ATP binding #status predicted <AT4>

Query Match 74.4%; Score 32; DB 1; Length 2330;
Best Local Similarity 62.5%; Pred. No. 8.9e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 SLLMWITQ 8
              | |:|:||
Db       1249 SRLLWVTQ 1256
```

RESULT 16
S44054
genome polyprotein - Marburg virus (strain Popp)
N;Alternate names: structural protein L
N;Contains: RNA-directed RNA polymerase (EC 2.7.7.48)

```
C;Species: Marburg virus
A;Variety: strain Popp
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 18-Jun-1999
C;Accession: S44054; S32776
R;Bukreyev, A.A.; Volchkov, V.E.; Blinov, V.M.; Dryga, S.A.; Netesov, S.V.
submitted to the EMBL Data Library, January 1994
A;Description: Full-length nucleotide sequence of Marburg virus Popp strain: The compari
A;Reference number: S44049
A;Accession: S44054
A;Molecule type: genomic RNA
A;Residues: 1-2331 <BUK>
A;Cross-references: EMBL:Z29337; NID:g450908; PIDN:CAA82542.1; PID:g450915
A;Experimental source: strain Popp
R;Bukreyev, A.A.; Netesov, S.V.
submitted to the EMBL Data Library, September 1992
A;Description: The partial nucleotide sequence of Marburg virus genome.
A;Reference number: S32775
A;Accession: S32776
A;Molecule type: genomic RNA
A;Residues: 1-2331 <BUW>
A;Cross-references: EMBL:X68494; NID:g296962; PIDN:CAA48508.1; PID:g296963
A;Experimental source: strain Popp
C;Genetics:
A;Gene: L
C;Superfamily: parainfluenza virus RNA-directed RNA polymerase
C;Keywords: ATP; nucleotidyltransferase

  Query Match             74.4%;  Score 32;  DB 2;  Length 2331;
  Best Local Similarity   62.5%;  Pred. No. 8.9e+02;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SLLMWITQ 8
              | |:|:||
Db       1249 SRLLWVTQ 1256


RESULT 17
E69786
ribosomal-protein-alanine N-acetyltransfer homolog ydiD - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C;Accession: E69786
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: E69786
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-151 <KUN>
A;Cross-references: GB:Z99107; GB:AL009126; NID:g2632866; PIDN:CAB12412.1; PID:g2632906
A;Experimental source: strain 168
C;Genetics:
A;Gene: ydiD
C;Superfamily: Escherichia coli ribosomal-protein-alanine N-acetyltransferase rimI

  Query Match             72.1%;  Score 31;  DB 2;  Length 151;
  Best Local Similarity   57.1%;  Pred. No. 95;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SLLMWIT 7
                :|:||:|
Db        142 ALIMWVT 148


RESULT 18
T07603
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) 22K chain - potato
C;Species: Solanum tuberosum (potato)
C;Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 03-Jun-2002
C;Accession: T07603
R;Heiser, V.; Grohmann, L.; Brennicke, A.
Plant Mol. Biol. 31, 1195-1204, 1996
A;Title: The plant mitochondrial 22 kDa (PSST) subunit of respiratory chain complex I i
A;Reference number: Z16044; MUID:97071689; PMID:8914535
A;Accession: T07603
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-213 <HEI>
A;Cross-references: EMBL:X96671; NID:g1235606; PIDN:CAA65451.1; PID:g1235607
A;Experimental source: cv. Desiree;green leaves
C;Genetics:
A;Gene: PSST
A;Genome: nuclear
C;Superfamily: psbG protein
C;Keywords: electron transfer; membrane-associated complex; mitochondrion; NAD; oxidore

  Query Match             72.1%;  Score 31;  DB 2;  Length 213;
  Best Local Similarity   85.7%;  Pred. No. 1.3e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LLMWITQ 8
              |||| ||
Db        207 LLMWWTQ 213


RESULT 19
S45677
proteinase inhibitor - signal crayfish
C;Species: Pacifastacus leniusculus (signal crayfish)
C;Date: 07-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 03-Feb-2003
C;Accession: S45677
R;Johansson, M.W.; Keyser, P.; Soederhaell, K.
Eur. J. Biochem. 223, 389-394, 1994
A;Title: Purification and cDNA cloning of a four-domain Kazal proteinase inhibitor from
A;Reference number: S45677; MUID:94333326; PMID:8055907
A;Accession: S45677
A;Molecule type: mRNA
A;Residues: 1-228 <JOH>
A;Cross-references: EMBL:X79512; NID:g498784; PIDN:CAA56043.1; PID:g498785
C;Keywords: serine proteinase inhibitor
F;21-66/Domain: Kazal proteinase inhibitor homology <KP1>
F;72-119/Domain: Kazal proteinase inhibitor homology <KP2>
F;123-170/Domain: Kazal proteinase inhibitor homology <KP3>
F;174-221/Domain: Kazal proteinase inhibitor homology <KP4>

  Query Match             72.1%;  Score 31;  DB 2;  Length 228;
  Best Local Similarity   85.7%;  Pred. No. 1.4e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SLLMWIT 7
              ||| |||
Db          4 SLLTWIT 10


RESULT 20
T11362
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - Daphnia pulex mitochondrion
C;Species: mitochondrion Daphnia pulex
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Jun-2002
C;Accession: T11362
R;Crease, T.J.
Gene 233, 89-99, 1999
A;Title: The complete sequence of the mitochondrial genome of Daphnia pulex (Cladocera:
```

A;Reference number: Z17264; MUID:99307147; PMID:10375625
A;Accession: T11362
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-311 <CRE>
A;Cross-references: EMBL:AF117817; NID:g4927669; PID:g4927682; PIDN:AAD33242.1
C;Genetics:
A;Gene: ND1
A;Genome: mitochondrion
A;Genetic code: SGC4
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 1
C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 72.1%; Score 31; DB 2; Length 311;
Best Local Similarity 71.4%; Pred. No. 1.9e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLMWIT 7
||||::
Db 83 SLLMWLS 89

RESULT 21
B83434
translocation protein in type III secretion PA1690 [imported] - Pseudomonas aeruginosa (
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: B83434
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathc
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: B83434
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-349 <STO>
A;Cross-references: GB:AE004596; GB:AE004091; NID:g9947658; PIDN:AAG05079.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: pscU; PA1690
C;Superfamily: flagellar biosynthetic protein flhB; flhB carboxyl-terminal homology

Query Match 72.1%; Score 31; DB 2; Length 349;
Best Local Similarity 71.4%; Pred. No. 2.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLMWIT 7
|||:|:|
Db 151 SLLVWLT 157

RESULT 22
C49274
protein farnesyltransferase (EC 2.5.1.-) beta subunit - bovine
N;Alternate names: farnesyl-protein transferase beta subunit; FTPase beta subunit; preny
C;Species: Bos primigenius taurus (cattle)
C;Date: 25-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 20-Sep-1999
C;Accession: C49274
R;Omer, C.A.; Kral, A.M.; Diehl, R.E.; Prendergast, G.C.; Powers, S.; Allen, C.M.; Gibbs
Biochemistry 32, 5167-5176, 1993
A;Title: Characterization of recombinant human farnesyl-protein transferase: cloning, ex
A;Reference number: A49274; MUID:93264431; PMID:8494894
A;Accession: C49274
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-437 <OME>
A;Experimental source: brain
A;Note: sequence extracted from NCBI backbone (NCBIP:132834)
C;Superfamily: DPR1 protein
C;Keywords: transferase

Query Match 72.1%; Score 31; DB 2; Length 437;
Best Local Similarity 71.4%; Pred. No. 2.7e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SLLMWIT 7
|||:|:|
Db 272 SLLQWVT 278

RESULT 23
B49274
protein farnesyltransferase (EC 2.5.1.-) beta subunit - human
N;Alternate names: farnesyl-protein transferase beta subunit; FTPase beta subunit; preny
C;Species: Homo sapiens (man)
C;Date: 25-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 01-Dec-2000
C;Accession: B49274; I60951
R;Omer, C.A.; Kral, A.M.; Diehl, R.E.; Prendergast, G.C.; Powers, S.; Allen, C.M.; Gibbs
Biochemistry 32, 5167-5176, 1993
A;Title: Characterization of recombinant human farnesyl-protein transferase: cloning, ex
A;Reference number: A49274; MUID:93264431; PMID:8494894
A;Accession: B49274
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-437 <OME>
A;Experimental source: placenta
A;Note: sequence extracted from NCBI backbone (NCBIP:132829)
R;Andres, D.A.; Milatovich, A.; Ozcelik, T.; Wenzlau, J.M.; Brown, M.S.; Goldstein, J.L.
Genomics 18, 105-112, 1993
A;Title: cDNA cloning of the two subunits of human CAAX farnesyltransferase and chromoso
A;Reference number: A47659; MUID:94102736; PMID:8276393
A;Accession: I60951
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 51-282,'L',284-437 <RES>
A;Cross-references: GB:L10414; NID:g388757; PIDN:AAA86286.1; PID:g388758
C;Genetics:
A;Gene: GDB:FNTB
A;Cross-references: GDB:138174; OMIM:134636
A;Map position: 14q23-14q24
C;Superfamily: DPR1 protein
C;Keywords: transferase

Query Match 72.1%; Score 31; DB 2; Length 437;
Best Local Similarity 71.4%; Pred. No. 2.7e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SLLMWIT 7
|||:|:|
Db 272 SLLQWVT 278

RESULT 24
A40037
protein farnesyltransferase (EC 2.5.1.-) beta chain - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 16-Oct-1992 #sequence_revision 16-Oct-1992 #text_change 05-Nov-1999
C;Accession: A40037
R;Chen, W.J.; Andres, D.A.; Goldstein, J.L.; Russell, D.W.; Brown, M.S.
Cell 66, 327-334, 1991
A;Title: cDNA cloning and expression of the peptide-binding beta subunit of rat p21(ras)
A;Reference number: A40037; MUID:91309145; PMID:1855253
A;Accession: A40037
A;Molecule type: mRNA
A;Residues: 1-437 <CHE>
A;Cross-references: GB:M69056; NID:g204185; PIDN:AAA41176.1; PID:g204186
C;Comment: This protein attaches farnesyl residues to a cysteine near the carboxyl termi
C;Superfamily: DPR1 protein
C;Keywords: heterodimer; transferase

Query Match 72.1%; Score 31; DB 2; Length 437;
Best Local Similarity 71.4%; Pred. No. 2.7e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 SLLMWIT 7
              ||| |:|
Db        272 SLLQWVT 278


RESULT 25
H96603
unknown protein F14G9.16 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: H96603
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H96603
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-752 <STO>
A;Cross-references: GB:AE005173; NID:g11094723; PIDN:AAG29658.1; GSPDB:GN00141
C;Genetics:
A;Gene: F14G9.16
A;Map position: 1

  Query Match          72.1%;  Score 31;  DB 2;  Length 752;
  Best Local Similarity   83.3%;  Pred. No. 4.5e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LLMWIT 7
              ||||:|
Db        223 LLMWLT 228


Search completed: August 22, 2004, 11:03:08
Job time : 40 secs
```

This Page Blank (uspto)